

nature



THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Invertebrate origins meet vertebrate-like complexity in the octopus genome **PAGE 220**

WITHIN ARMS' REACH

SOCIAL POLICY

A CHANCE TO SUCCEED

'Randomistas' put anti-poverty interventions on trial

PAGE 150

SUSTAINABILITY

TOO GOOD TO WASTE

Mining shellfish waste for chemical riches

PAGE 155

CAREERS

THE RUSH TO BIOLOGICS

Investment in biopharma a boost for young researchers

PAGE 257

NATUREASIA.COM

13 August 2015

Vol. 524, No. 7564

THIS WEEK

EDITORIALS

WORLD VIEW Europe is wrong to turn its back on Science 2.0 **p.137**

NEPAL Trio of studies shed light on Kathmandu earthquake **p.139**



OBSCURA New image shows Earth and Moon's dark side **p.140**

Science in the community

Randomized controlled trials are providing evidence about interventions in health, education and international development, but they are only part of a suite of useful tools.

The past decade has seen an explosion of interest in the use of randomized controlled experiments to test public policies, on issues from health and public safety to agriculture and education. Practitioners are generating hard data and pushing evidence into the government sphere. But they are ruffling the feathers of conventional economists, who have long focused on models and qualitative field data. Some worry that the new focus on randomization could skew the questions that researchers look at, or could produce black-box results that cannot answer crucial questions about why something does or does not work. But more evidence is always a good thing, and there is plenty of room for academics of all stripes.

Despite disagreements, it is hard to dispute the value of the basic goal: to ensure that governments invest their limited resources in programmes that work as advertised, and look for ways to alter or eliminate those that do not. But in the messy space occupied by social scientists, it is not always easy to determine which programmes work, which do not, and why. Enter the randomized controlled trial, in which changes are measured in a selection of individuals or groups who have been randomly assigned to receive an intervention — or not. The medical industry has used such trials to tease out the effects of drugs for decades.

As discussed on page 150, development economists have led the way, and are now running hundreds of trials that are designed to improve the effectiveness of international aid and, ultimately, the much larger pool of domestic spending by governments in developing countries. These researchers have already produced valuable insights, and governments are scaling up some of the results. Buses in Kenya are becoming safer thanks to stickers that urge passengers to speak up when they feel unsafe, and residents of the Indian state of Gujarat may soon benefit from the implementation of a new pollution-auditing system for industrial plants.

EVIDENCE-BASED INTERVENTION

Many of the trials focus on human behaviour in very particular circumstances, but researchers are targeting larger questions as well. Some studies have looked at the impact of community-driven development efforts, with mixed results. Others are considering how trials could provide the evidence that governments and aid agencies need to improve the delivery of humanitarian relief. And then there are the long-standing questions about how much aid can accomplish in lifting people out of poverty. In a study published in May, a team of researchers ran randomized trials in six countries looking at whether a package of interventions that included cash, food, health care and training could give a lasting boost to the poor (A. Banerjee *et al.* *Science* <http://doi.org/4p7>; 2015). The evidence suggests that the answer is yes, for at least one year after the intervention has ceased (see *Nature* **521**, 269; 2015).

Contrast that with the Millennium Villages Project (MVP), which

began in 2004 and has pushed a comprehensive aid package into villages in ten countries in Africa. MVP researchers have begun evaluating how the project has done so far (see page 144). But the research protocol that they published last month (see go.nature.com/3eidfr) acknowledges that it will be difficult to definitively answer questions about how well these villages have fared compared with surrounding villages that did not receive the intervention, largely because the project did not use an experimental approach from the outset. Information on whether the

project is effective would have been useful for policymakers who are facing difficult choices about where to invest limited resources.

“The goal is to ensure that governments invest in programmes that work.”

Many poverty-alleviation programmes focus on providing money with strings attached — only if families keep their children in school and attend health clinics, for instance — but some researchers are now

advocating unconditional cash transfers. Paul Niehaus, an economist at the University of California, San Diego, co-founded the non-profit organization GiveDirectly in New York City to do just that. His team points out that many forms of development aid are complex and costly to administer, and suggests judging their effectiveness against that of simply giving money. This would be done in randomized trials, in which the intervention group receives development aid and the control group gets cash.

There is much to be learned from randomized trials, but researchers must acknowledge that such experiments have limitations, not least that they do not necessarily provide answers as to why something works or does not. In the case of community-driven development programmes, for instance, a randomized controlled trial can provide basic statistics about whether letting community councils make their own decisions hastens the delivery of basic services, improves the economy and advances the social well-being of women. But do these councils actually promote social cohesion in a meaningful way? The results so far are mixed, and this is where agencies and institutions can benefit from the soft, qualitative social-science survey data that researchers have been seeking to go beyond.

Randomized controlled trials will not be able to provide answers to all of the world's enduring questions, and they are not the only way to gather solid data. Economists have developed other quasi-experimental approaches that do not include randomization but can nonetheless provide rigorous statistics with which to judge a programme's effectiveness.

In combination, these methods are providing policymakers with more information every day. But perhaps more important is what happens next. Politicians, government bureaucrats, activists and philanthropists will all be happy to talk about the value of randomized trials when the results support their policies and programmes. They must also have the courage to do so when the evidence goes the other way. ■



The future of science will soon be upon us

The European Commission has abandoned consideration of 'Science 2.0', finding it too ambitious. That was the wrong call, says Colin Macilwain.

As the staff of the European Commission head for the beaches this August, they have been asked to ponder the future of science. Research commissioner Carlos Moedas has announced his priorities as being “open science” and “open innovation”, and invited his team to report back with its ideas on how to achieve that.

These goals sound laudable enough, but they're really rather anodyne. Sadly, the commission has closed the door on a more ambitious project. This time last year, it sought views on Science 2.0. That term infers truly radical change — including rapid evolution of the two main pillars that underpin science: the research paper and the single-investigator grant.

Predictably, the first noises to emanate from scientific leaders to the Science 2.0 consultation were sighs of scepticism. Science, they purred, organizes itself indigenously from the bottom up. Each discipline has developed its own processes and the arrival of the digital economy — the main spur for a Science 2.0 reboot — was an incremental change that the disciplines will accommodate in their own ways, in the fullness of time.

I fully understand why no one wants the commission to get carried away and do something drastic. I can't help feeling, however, that an opportunity has been lost. A revamp of the commission's €11-billion (US\$12-billion)-a-year research programmes to anticipate Science 2.0 might have nurtured the ability of a new generation of Europeans to develop knowledge in different ways.

Take the peer-reviewed paper, the main yardstick for success or failure in almost all academic research careers. The paper — as its name handily implies — has been rendered obsolete by the arrival of the online world. The constraints that paper publication places on overlapping authorship, evolution of content and links to other work and to other people's data, have all gone.

A decade ago, many people in scientific publishing envisaged the broadening or abandonment of the discrete paper by around now. Yet because of the extent to which institutions, career paths, publishers and funding agencies rely on this essentially outdated concept, the paper stubbornly persists.

Indeed, the measurement of academic achievement in terms of papers published in highly cited journals, such as this one, has evolved into a kind of fetish. The 2012 San Francisco Declaration on Research Assessment, in which almost everyone in science declared their mutual abhorrence for reliance on citation data, stands, in truth, as testament to the manner in which such data now bestride the scientific world.

But the transformation of the paper is inevitable in the long run, as bulk data are shared, research outcomes can be continually updated

and people demand appropriate credit for their work.

The second pillar — the peer-reviewed, single-investigator grant — may hang around for longer. Since its development in the middle of the last century, after all, the distribution of such grants have evolved as the most tried-and-tested method of public support for science.

The difficulty here is growing political impatience with the promised outcomes from this grant funding. The annual budget of the US National Institutes of Health, the largest disburser of such grants in the world, has been frozen now for more than a decade at just over \$30 billion. As a result, it is increasingly older people, who know how to work the system, who get funding: people under 40 are finding it harder and harder to get their foot on the ladder. One consequence is the steady drumbeat from Congress for ‘prizes’ and other gimmicks to circumvent peer review.

The UK research councils are in similarly dire straits. The new secretary of state for business, Sajid Javid, who has been pressed to look for cuts of 25–40% in his budget, last month brought in the US consultancy firm McKinsey to look at how the research councils function. Bodies such as the UK Medical Research Council have a fine and well-deserved global reputation. But when McKinsey looks under the hood, it may well discover that the outputs are not what the politicians are after.

What politicians want, these days, is ‘innovation’: an odd hybrid, sitting somewhere between science, engineering, finance and human inclination. As everyone knows, ‘innovation’ thrives most visibly in places such as Silicon Valley. But would-be innovators there already occupy a culture in which spending several years on a PhD and then

grinding your way to a professorship simply isn't part of the currency. The lifestyle changes that are under way in innovators during their most creative years do not align with the decade-old funding approaches used today by the European Commission and other major funding agencies.

There is no easy way to alter the architecture of these funding routes. The divisional structure of the US National Science Foundation, for example, is recognized even by its most senior officials as inappropriate in today's multidisciplinary world. But they can't propose change, lest Congress makes a hash of it.

By contrast, the commission has some room for manoeuvre — and the luxury of long-term planning. The programme that Moedas is starting to prepare now will run from 2021 to 2027, by which time changes triggered by the Internet and globalization will surely be impossible to resist. The commission should be planning for how the world will look then — and envisioning Science 2.0 would be a good place to start. ■

Colin Macilwain writes about science policy from Edinburgh, UK.
e-mail: cfmworldview@googlemail.com

THE
COMMISSION
HAS SOME ROOM FOR
MANOEUVRE,
AND THE
LUXURY
OF LONG-TERM
PLANNING.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/7h3bkj

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

CONSERVATION

Poverty drives forest raiders

Researchers have proposed ways to improve efforts to stop people illegally harvesting wood and food from a conservation park in Uganda.

Bwindi Impenetrable National Park, which covers 330 square kilometres, hosts half of the world's endangered mountain gorillas. Mariel Harrison at Imperial College London and her colleagues surveyed 365 households around Bwindi and found that people in 26% of them had hunted for bushmeat and 20% had collected firewood in the park. These illegal activities were most prevalent among the poorest households living in remote areas and closest to the park. Focus groups reported poverty and resentment at the inequitable distribution of benefits from the park as reasons for their activities.

The researchers suggest that projects combining conservation and development should benefit the poorest people in remote areas near park boundaries to reduce illegal activities that jeopardize the park's resources.

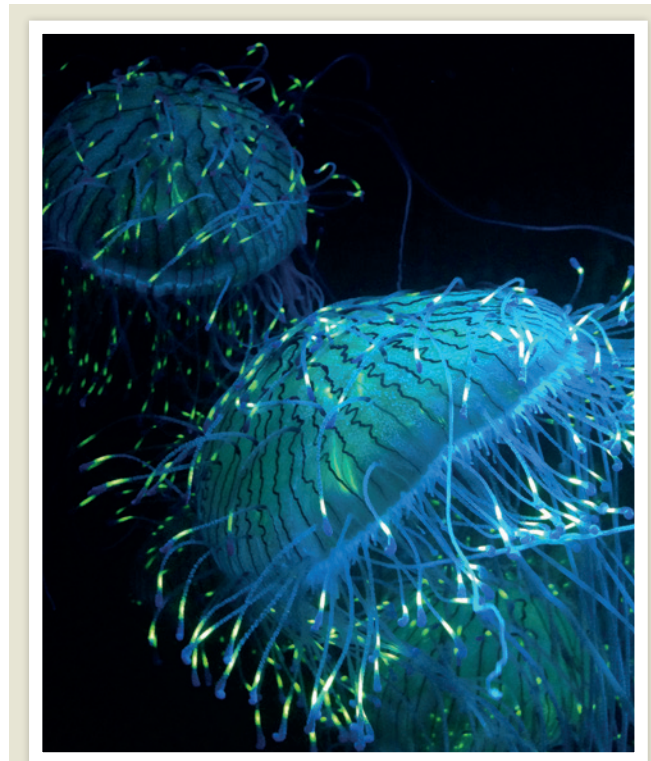
Conserv. Biol. <http://doi.org/6nt> (2015)

NEURODEVELOPMENT

Mouse brain cells made primate-like

By turning on a single gene in specific neural cells in the embryonic mouse brain, researchers have made more neurons grow in the neocortex — a region that evolved to be much larger in primates than in other mammals.

Wieland Huttner at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany, and his



ZOOLOGY

Light show lures prey

Jellyfish and other marine animals could be using their fluorescent proteins to attract prey.

Proteins such as green fluorescent protein (GFP) are invaluable tools for biologists in the lab, but their role in nature has not been clear. Steven Haddock at Monterey Bay Aquarium Research Institute in Moss Landing, California, and Casey Dunn at Brown University in Providence, Rhode Island, placed the flower hat jellyfish (*Olindias formosus*, pictured in blue light) in a tank along with its rockfish prey and separated the two with a transparent wall. When they exposed *O. formosus* to blue light (the light of its underwater habitat), the tips of its tentacles fluoresced green and the rockfish attacked the barrier more often than under yellow or white light or when the jellyfish was replaced with a non-fluorescent decoy.

Herbivorous prey may be seeing the fluorescence as an indicator of chlorophyll, which also fluoresces.

Biol. Open <http://doi.org/6qd> (2015)

team developed a mouse model in which they could switch on the *Pax6* gene in specific neural progenitor cells — where it is expressed in humans but not in mice. They turned on the gene in cells that give rise

to neurons of the neocortex, which controls advanced cognitive abilities. The team found that with sustained *Pax6* expression, the progenitor cells proliferated more, resulting in more neurons in

parts of the neocortex.

Pax6 could have had an important role in the evolution of the larger primate neocortex, the authors suggest. *PLoS Biol.* 13, e1002217 (2015)

ASTRONOMY

Stars align to show new planet

Two teams using different telescopes have confirmed that a planet with a mass similar to that of Uranus is orbiting a distant star.

Most known exoplanets orbit close to their stars, but in 2005 researchers using an effect called microlensing spotted a planet with a larger orbit. This effect happens when two stars align: the gravity of the star in front magnifies light from the one behind. Planets in the foreground system can alter this light, which allows them to be detected.

David Bennett at the University of Notre Dame in Indiana and his colleagues used NASA's Hubble Space Telescope to study the light from the microlensing event OGLE-2005-BLG-169 more precisely. Their observations indicated the presence of a planet roughly 14 times heavier than Earth and more than 3 times farther from its star.

Another team that included Bennett, led by Virginie Batista of the Paris Institute of Astrophysics, used the W. M. Keck Observatory in Hawaii and found similar properties for the planet.

Astrophys. J. 808, 169;

Astrophys. J. 808, 170 (2015)

VACCINOLOGY

How the Ebola vaccine protects

The Ebola vaccine that proved effective in a trial of more than 4,000 people in Guinea seems

S. HADDOCK/MBARI/HTTP://BIOLUN.EEMBUCS.EDU

to work by rapidly triggering one arm of the immune system to hold back the virus while the body ramps up antibody production, according to a study in monkeys.

Heinz Feldmann of the National Institute of Allergy and Infectious Diseases in Hamilton, Montana, and his colleagues tested the VSV-EBOV vaccine, which was designed to fight the 2014 West African outbreak strain of Ebola virus. The team immunized 15 rhesus macaques (*Macaca mulatta*) and then infected them with the virus. All but one of the vaccinated animals survived, whereas all unimmunized animals died about a week after infection.

Analysis of the surviving animals' blood showed that the vaccine triggered the innate immune system to keep viral replication in check during the first days of infection, giving the rest of the immune system time to churn out Ebola-specific antibodies.

Science <http://doi.org/6p9> (2015)

GEOLOGY

Why Nepal quake was so damaging

The magnitude-7.8 earthquake that devastated much of Nepal on 25 April did not relieve all of the geological stress in the region — making another big quake probable.

A team led by Jean-Philippe Avouac at the University of

Cambridge, UK, used seismic data and satellite radar to show that a 140-kilometre stretch of a major Himalayan geological fault shifted during the disaster. This transferred stress into neighbouring areas of the fault, which may now be more prone to rupturing in a future quake.

In a separate paper, Yuji Yagi and Ryo Okuwaki of the University of Tsukuba, Japan, found that the earthquake rupture raced eastward from its point of origin. The greatest movement of the fault occurred about 50 kilometres east of the quake's epicentre — close to Kathmandu.

This discovery helps to explain why the shaking was so destructive to the city, according to a third paper by another team led by Avouac. Although much of the ground in the region shook only moderately, the seismic energy was amplified across the Kathmandu basin in ways that caused tall buildings, including temples, to sway and collapse (pictured is what remains of Kathmandu's Bhimsen Tower, known as Dharahara).

Nature Geosci. <http://doi.org/6p7> (2015); **Geophys. Res. Lett.** <http://doi.org/6ns> (2015); **Science** <http://doi.org/6p6> (2015)

CELL BIOLOGY

Chemicals switch cells' identity

Adult skin cells have been transformed directly into neurons by two independent groups in China using just small-molecule chemicals.

Reprogramming adult cells back into stem cells or directly into other types of specialized cells requires transcription factors, which modify cells genetically. To avoid tinkering with the cells' genes, Gang Pei and Jian Zhao from the Shanghai Institutes for Biological Sciences and their colleagues worked with fibroblasts, or skin cells, from both healthy adults and people with Alzheimer's disease, culturing them with a cocktail

SOCIAL SELECTION

Popular topics on social media

Bioethics comes under fire

The latest biomedical technologies, from fetal stem cells to human gene editing, offer huge potential for treating disease. They also raise tricky ethical questions that can eventually result in guidelines on how to prevent their misuse. In an opinion piece in *The Boston Globe*, Harvard University psychologist Steven Pinker argues that this sweeping ethical oversight delays innovation and should 'get out of the way' (go.nature.com/93t5ti). The article ignited much discussion on social media among bioethicists and researchers. Many disagreed with Pinker, including Daniel Sokol, a London-based bioethicist and lawyer, who wrote in a blog post that ethicists should at times 'get in the way' (see go.nature.com/zmluki). Research to alleviate human suffering is important, he added, but "misguided attempts to help can — and have — led to incalculable harm".

NATURE.COM
For more on popular papers:
go.nature.com/b6jxzu

of small molecules to produce neurons.

Hongkui Deng at Peking University, Beijing, and his colleagues used a different set of chemicals to convert mouse fibroblasts into neurons. Both groups made neurons that looked, fired and made functional connections just like neurons created from fibroblasts using transcription factors.

The chemicals modulate key molecular signalling pathways to change an adult cell's identity. This approach could make it easier to reprogram cells for clinical use, say the authors.

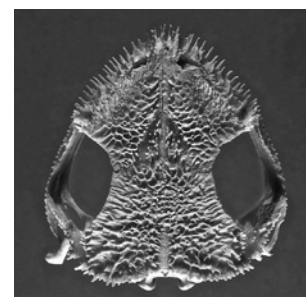
Cell Stem Cell <http://doi.org/6p4>; **Cell Stem Cell** <http://doi.org/6p5> (2015)

ZOOLOGY

Venomous frogs headbutt foe

Two Brazilian frog species use sharp spines protruding from around their noses and mouths to deliver toxins in their skin to predators — the first evidence of a venomous frog.

Most frogs produce toxins in their skin but have no way of deliberately passing them on to predators. Edmund Brodie at Utah State University



in Logan and his team discovered this toxin delivery when they were collecting specimens of two tree-frog species (*Corythomantis greeningi* and *Aparasphenodon brunoi*) and restraining them in their hands. *C. greeningi* (skull, pictured) jabbed its spiny head into the offending hand and released toxins from its skin glands, causing intense pain in the arm for several hours. When tested in mice, the venom from both frog species caused swelling and was deadly at high concentrations.

There could be more venomous amphibians than thought, the authors say. **Curr. Biol.** <http://doi.org/6n7> (2015)

NATURE.COM
For the latest research published by Nature visit:
www.nature.com/latestresearch



SEVEN DAYS

The news in brief

FACILITIES

Virus lab for Japan

Japan has cleared the way for its first lab to handle the highest-risk pathogens, such as Ebola virus, at a facility about 30 kilometres west of Tokyo. The National Institute of Infectious Diseases originally built the facility in Musashi-Murayama in 1981 to operate at the greatest biosafety level, BSL-4. But opposition from local residents has forced it to run as a BSL-3 lab. On 3 August, the country's health minister reached an agreement with the mayor of Musashi-Murayama that would allow BSL-4 operations to begin.

PEOPLE

Energy director

US President Barack Obama will nominate physicist Cherry Murray as the director of the US Department of Energy's Office of Science, the White House announced on 5 August. The position has been vacant since 2013. Murray, now at Harvard University in Cambridge, Massachusetts, was principal associate director for science and technology at California's Lawrence Livermore National Laboratory from 2007 to 2009. She also worked at Bell Laboratories in Murray Hill, New Jersey, from 1978 to 2004. At the energy department, Murray will oversee a US\$5-billion research budget.

POLICY

Interrogation ban

The American Psychological Association (APA) has decided to ban psychologists from participating in military interrogations, responding to a damning report implicating the organization in the torture of detainees by US military and intelligence agencies. At the

organization's annual meeting in Toronto on 7 August, APA's council of representatives voted to approve the measure by 156 to 1, with 7 abstentions. The APA parted company with several senior officials after the report was released, and it says that it will convene a panel to review its ethics guidelines.

Scots say no to GM

Scotland has declared its intention to opt out of growing genetically modified (GM) crops approved by

the European Union (EU) on non-scientific grounds — the first region to do so under new EU rules. Rural affairs secretary Richard Lochhead said on 8 August that Scotland wished to protect its 'clean, green' status. The rules were introduced in April to overcome a political impasse in which EU member states that are divided on the principle of GM crops have blocked approvals for safety-cleared crops. Countries have until 3 October to opt out of

the varieties currently being assessed by the European Commission for cultivation.

Australia emissions

Australia's government plans to lower national greenhouse-gas emissions by 26–28% below 2005 levels by 2030, it announced on 11 August. It is the latest country to state its commitment ahead of a United Nations global-warming summit in Paris this December. Prime Minister Tony Abbott said that the



NASA/NOAA

Moon snapped from its dark side

NASA's DSCOVR satellite captured this image of the Moon crossing in front of Earth from a vantage point 1.6 million kilometres away. DSCOVR monitors space weather from the gravitationally stable L1 point. The far side of the Moon has fewer dark lava flows — or maria

— than the side that faces Earth. The most prominent flow (dark blotch, upper left) is Mare Moscoviense, named by the Soviet Union after its probe Luna 3 took the first photos of the far side of the Moon in 1959. The DSCOVR photo was released on 5 August.

LUCAS MARIE/AP

targets balanced the need to slow climate change and to promote strong economic growth. Australia has the highest per capita emissions of any of the 34 industrialized countries in the Organisation for Economic Co-operation and Development.

EVENTS

End for Ada project

The Ada Initiative, which has spent four years addressing sexism and harassment at science and technology conferences, announced on 4 August that it will cease activities in mid-October. The non-profit organization has sought to eliminate environments that discriminate against women. It has encouraged conferences in fields including artificial life, physics and entomology to formally adopt anti-harassment policies. The organization said it decided to close after it had difficulty finding a suitable head to replace its founders.

Debris discovery

A section of an aircraft wing, found on an island in the Indian Ocean on 29 July, is probably from missing Malaysia Airlines flight MH370, a French prosecutor said on 5 August. The announcement came after investigators in Toulouse, France, examined the debris



(pictured) and determined that it is from a Boeing 777 — the same model as the missing plane. In a stronger statement, Malaysian Prime Minister Najib Razak announced definitively that the wing fragment belonged to MH370. The families of passengers who were on the flight reacted to the mixed messages with disbelief and anger.

Iran deal backed

A group of 29 leading US physicists and experts on nuclear weapons and arms control wrote a letter to President Barack Obama on 8 August supporting the Iran nuclear agreement made with the United States and five other countries on 14 July. The letter's authors include the chief executive of the American Association for the Advancement of Science; a former head of the Los Alamos National Laboratory in New Mexico; and six Nobel prizewinners. They say that the deal includes "more stringent

constraints than any previously negotiated non-proliferation framework". The letter is intended to sway undecided US Congress members ahead of a vote that could derail the agreement in September.

FUNDING

Ticket to space

NASA has purchased rides worth US\$490 million aboard Russia's Soyuz spaceships, agency head Charles Bolden wrote to the US Congress on 5 August. Ever since the US space shuttle programme ended in 2011, NASA has relied on Russia to transport its astronauts to and from the International Space Station. The contract extends their agreement to the end of 2018, adding to the \$458 million that NASA is already paying for Soyuz flights in 2017. The agency is seeking private ways to ferry astronauts to space, but cuts to government funding of the 'commercial-crew' programme have hindered progress.

BUSINESS

Rare diseases bid

Drug firm Shire has proposed a US\$30-billion hostile takeover of its competitor Baxalta in Bannockburn, Illinois. Both companies make drugs to treat rare diseases. The offer from Dublin-based Shire was made

COMING UP

13 AUGUST

The Rosetta spacecraft, which is studying the comet 67P/Churyumov-Gerasimenko in detail, reaches perihelion — the closest point to the Sun during its journey.

16–20 AUGUST

Around 12,500 chemists gather in Boston, Massachusetts, for the 250th American Chemical Society meeting.

go.nature.com/5b3gxt

16–21 AUGUST

Geochemists meet at the 25th Goldschmidt conference in Prague. To mark the anniversary, 25 lectures will highlight geochemical advances since the first conference.

go.nature.com/hu9x5x

public on 4 August to appeal directly to Baxalta shareholders after an initial approach was rebuffed privately on 10 July. Baxalta's directors say that the offer significantly undervalues the company, which specializes in treatments for immune disorders and rare blood conditions.

3D-printed pills

The US Food and Drug Administration (FDA) approved the first 3D-printed drug, an antiepileptic medication called levetiracetam, on 3 August. Manufactured by Aprelia Pharmaceuticals in Langhorne, Pennsylvania, the printed drug is a porous pill that dissolves in the mouth to make it easy to swallow. The company expects the drug to arrive on the market in early 2016 and intends to produce more 3D-printed therapies for central nervous system disorders.

➔ NATURE.COM

For daily news updates see:

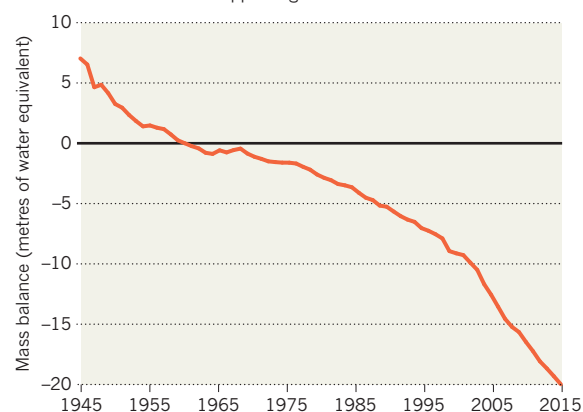
www.nature.com/news

TREND WATCH

The current rate of global glacier retreat is "historically unprecedented", researchers at the World Glacier Monitoring Service reported on 30 July (M. Zemp *et al.* *J. Glaciol.* **61**, 745–762; 2015). The team compiled more than 47,000 observations, dating as far back as the sixteenth century, and found that glaciers are now shrinking nearly twice as fast as they were during the late twentieth century. Even if the climate stabilizes, ice loss is expected to continue in many areas.

GLACIERS RETREATING EVER FASTER

The world's glaciers have been shrinking since the 1960s, but their cumulative mass is now disappearing faster than ever before.



WORLD GLACIER MONITORING SERVICE

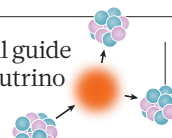
NEWS IN FOCUS

DEVELOPMENT Star-studded aid effort launches bid to measure impact **p.144**

BIOTECHNOLOGY GM-crop activists target scientists' e-mails **p.145**

PHYSICS Graphical guide to next era of neutrino discovery **p.148**

DEVELOPMENT Randomized controlled trials find a role in social policy **p.150**



ASAHI SHIMBU/GETTY



Workers at the Sendai nuclear power plant conduct an emergency safety drill ahead of the restart that ended Japan's two-year nuclear freeze.

ENERGY POLICY

Japan ends nuclear hiatus

Return to nuclear energy will reduce carbon emissions but not by nearly enough.

BY DAVIDE CASTELVECCHI

The Sendai Nuclear Power Plant on the island of Kyushu broke a four-year lull on 11 August when it switched one of its reactors back on. The restart is the first since Japan's nuclear-power industry ground to a halt two years ago following safety concerns in the wake of the 2011 Fukushima Daiichi disaster.

It will help the world's third-largest economy to lower its carbon emissions. But the government energy plan that includes this shift in policy is much too modest if Japan is to help keep global temperatures from rising by more than 2°C above pre-industrial levels, say analysts.

The plan essentially returns the nation to its pre-Fukushima energy mix of mainly coal

and nuclear power, apart from a small but substantial increase in solar. "The mindset of the government and the heavy industry is still the same: to try to keep nuclear and also coal," says Tetsunari Iida, head of the Institute for Sustainable Energy Policies in Tokyo.

Before the disaster, Japan had aimed to produce about half of its electricity from nuclear sources. Following the meltdowns, the short-lived cabinet of then prime minister Yoshihiko Noda considered phasing out nuclear energy entirely (see *Nature* **486**, 15; 2012) and replacing it with renewables and fossil fuels.

However, the current government of Prime Minister Shinzo Abe, which took over in 2012, has put nuclear back into the picture, with plans to restart as many reactors as possible (see *Nature* **507**, 16–17; 2014). In July, the

government submitted its targets for reducing greenhouse-gas emissions to the United Nations, ahead of a UN global-warming summit in Paris this December. The pledge included a goal for nuclear energy to fulfil at least 20% of Japan's electricity needs by 2030. Renewable sources — mostly hydropower but also solar — would contribute a minimum of 22%.

This would reduce Japan's carbon footprint compared with the years since Fukushima, when electricity companies bridged the nuclear gap by ramping up the use of coal, oil and, especially, liquefied natural gas. But fossil fuels would still account for more than half the power generated in 2030. Nuclear and renewables would help keep carbon dioxide emissions in check, but overall emissions would be cut by only 18% from 1990 levels. The European ►

► Union, by comparison, pledged 40% cuts from 1990. “I think that the government understands and acknowledges the climate goal and tries to make its target consistent with it, but industrial and economic criteria such as lowering electricity costs are given higher priority,” says Seita Emori, who heads a climate risk-assessment team at Japan’s National Institute for Environmental Studies in Tsukuba. The 2030 emissions target “doesn’t look really sufficient for the climate goal”.

The government sees an especially modest role for wind, projected to contribute only 1.7% of electricity generation by 2030. (Germany, by comparison, already derives around 8–9% of its power from wind.) Iida says there is an “irrational bias” against wind that is deep-rooted in Japan’s energy industry.

Moreover, the way Japan’s energy market

is structured, with a few de facto regional monopolies, is stacked against wind, favouring instead sources that are established, such as nuclear and fossil fuels. “Power companies control both the grid and existing power plants,” says Tomas Kåberger, head of the Tokyo-based Japan Renewable Energy Foundation. Wind would take a share of the market away from the utilities’ power plants, but the same utilities could deny wind-power companies access to the grid, says Ali Izadi-Najafabadi, who heads the Tokyo office of the consulting company Bloomberg New Energy Finance. The utilities must cite “technical grounds” for such a refusal, but “there is no independent grid operator, so it’s hard to judge those technical grounds,” he says.

To switch back on, the Sendai plant had to satisfy increased scrutiny from regulators and

the courts. Following the 2011 meltdowns, the Japanese government overhauled its nuclear safety policy, reviewed its atomic-energy infrastructure and created the independent Nuclear Regulation Authority (NRA). Izadi-Najafabadi says that the NRA showed that it can bite as well as bark when it forced utilities to decommission some of their more troublesome reactors. Still, anti-nuclear advocates complain that reforms have not gone far enough, in particular on evaluating seismic and volcanic risks and preparing evacuation plans, and that the NRA has bowed to political pressure to speed up its reviews.

Nuclear-safety culture has made progress since Fukushima, says Amory Lovins, co-founder of the Rocky Mountain Institute, an energy think tank in Snowmass, Colorado. But, he adds, “there is still a troublesome and pervasive lack of transparency”. ■

DEVELOPMENT

Flagship aid programme up for evaluation

The Millennium Villages Project in Africa begins analysis of first ten years to test impact.

BY JEFF TOLLEFSON

The Millennium Villages Project (MVP) stands out among development efforts in Africa. Since its launch in 2004, it has attracted generous donations and high-wattage supporters — including Hollywood actor Angelina Jolie and United Nations secretary-general Ban Ki-moon — for its work on alleviating poverty in rural Africa. The programme has delivered aid to at least 500,000 people in 10 countries, and has been emulated in others.

But its effectiveness has never been thoroughly tested. With the publication of a research plan in *The Lancet* last month (see go.nature.com/3eidfr), the MVP has now embarked on its first comprehensive evaluation in the hope of addressing long-standing questions about its impact.

Led jointly by the Earth Institute at Columbia University and the non-profit organization Millennium Promise, both in New York, the MVP aims to lift clusters of villages out of poverty through interventions ranging from building health centres, roads and schools to improving agriculture and sanitation. It began in Kenya in 2004, and as of 2013 had an annual budget of about US\$25 million.

“We will, I believe, be able to gain very useful insights about costs, processes, technologies, information systems, local and national



Building water-delivery systems is one measure used by the Millennium Villages Project in Africa.

governance, among other issues,” says Jeffrey Sachs, who runs the Earth Institute and who conceived the programme.

As economists increasingly advocate randomized controlled trials of international aid programmes (see page 150), Sachs has faced criticism for not setting up the MVP as a rigorous experiment. MVP researchers are now trying retroactively to compare villages that received the full intervention with similar ones

that did not, but the research protocol readily acknowledges challenges in collecting data and producing statistically significant results.

“I expect that the authors will conclude that, although we cannot prove that MVP works, we also cannot rule out that it works,” says Annette Brown, who heads the Washington DC office of the International Initiative for Impact Evaluation, a non-profit organization that funds and analyses such evaluations.

GUILLAUME BONNI/CORBIS

BIOTECHNOLOGY

Anti-GM group expands probe into industry ties

Activists seek release of records from 40 researchers at US public universities.

BY KEITH KLOOR

In March, preliminary results from a study commissioned by the UK Department for International Development (DFID) in 2011 found little benefit from an £11.5-million (\$18-million) expansion of the Millennium Villages project in northern Ghana. Sachs asserts that effects are hard to see at the Ghana site because it is in its early stages; his critics see the analysis as further evidence that the Millennium Villages approach may not work as advertised. “The trumpeting of the project as a model is just indescribably disproportionate to the deafening silence about its actual results,” says Michael Clemens, a senior fellow at the Center for Global Development, a non-profit think tank in Washington DC.

Clemens has long been an outspoken critic of the MVP and was among researchers who challenged¹ a 2012 study in *The Lancet*² that reported that child mortality had dropped in Millennium Villages three times faster than elsewhere in the host nations. The challenge ultimately led to a retraction of that claim by the paper’s lead author. Clemens argues that aid money should be spent either on projects that generate useful knowledge or on things that have been shown to work, noting that malaria bednets, which have a demonstrated benefit and are part of the MVP’s suite of interventions, cost \$15–20 per household.

The MVP typically budgets \$120 per capita annually, according to its website, although Sachs says that outside contributions can reduce MVP’s investment to half that. At the Ghana site funded by DFID, the total investment by all parties was projected to be \$27.1 million over 5 years for 30,000 people. That is \$181 per person annually, or about \$4,500 per household over the course of the project — less than the \$5,408 per household calculated by a randomized controlled trial in Ghana testing a two-year package of interventions that included food, cash, health services and training³.

The MVP hopes to release its analysis by the end of 2016, and Sachs says that his team will be in a better position to talk about cost-effectiveness and other considerations once the analysis is out.

Dean Karlan, an economist at Yale University in New Haven, Connecticut, says that it is probably too late for the project itself to advance the science of global development in a significant way, but he credits Sachs with raising awareness about global poverty issues. “I do see it as a missed opportunity,” Karlan says, “but in the grand scheme of things there are tons of missed opportunities.” ■ **SEE EDITORIAL P.135**

Michelle McGuire, a nutrition scientist at Washington State University in Pullman, was stunned last month when activists who oppose the use of genetically modified (GM) organisms asked to read her e-mail.

US Right to Know of Oakland, California, filed a request under Washington’s freedom-of-information law to see her correspondence with, or about, 36 organizations and companies. McGuire is one of 40 US researchers who have now been targeted by the group, which is probing what it sees as collusion between the agricultural biotechnology industry and academics who study science, economics and communication.

That investigation, which began in February, has just started to yield documents. These include roughly 4,600 pages of e-mails and other records from Kevin Folta, a plant scientist at the University of Florida in Gainesville and a well-known advocate of GM organisms. The records, which the university gave to US Right to Know last month, do not suggest scientific misconduct or wrongdoing by Folta. But they do reveal his close ties to the agriculture giant Monsanto, of St Louis, Missouri, and other biotechnology-industry interests.

The documents show that Monsanto reimbursed Folta for trips he took to speak to US students, farmers, politicians and the media. Other industry contacts occasionally sent him suggested responses to common questions about GM organisms.

“Nobody ever told me what to say,” says Folta, who considers public outreach to be a key part of his job. “There’s nothing I have ever said or done that is not consistent with the science.”

He adds that he has never accepted honoraria for outreach work, and that the University of Florida does not require him to disclose travel reimbursements. But the e-mails show that Folta did receive an unrestricted US\$25,000 grant last year from Monsanto, which noted that the money “may be

used at your discretion in support of your research and outreach projects”. Folta says that the funds are earmarked for a proposed University of Florida programme on communicating biotechnology.

Monsanto spokeswoman Charla Lord says that the company was “happy to support Dr Folta’s proposal for an outreach programme to increase understanding of biotechnology”, and that the \$25,000 grant “predominately covered travel expenses”. Lord adds that Monsanto considers public-private collaborations to be “essential to the advancement of science”.

Such explanations do not satisfy Gary Ruskin, executive director of US Right to Know. “I think it’s important for professors who take money from industry to disclose it,” he says. “And if they’re not disclosing it, that’s a problem. And if they say they aren’t taking money, and they are, then that’s a problem.”

Ruskin’s group, which was founded in 2014, calls for mandatory labelling of food that contains GM ingredients — even though numerous scientific bodies, including the US National Academy of Sciences, have found no evidence that such food harms human health.

US Right to Know launched its investigation of academic researchers after it noticed that several had fielded questions about crop biotechnology on a website called GMO Answers, which is funded by members of the biotech industry. The group considers the site, which is aimed at consumers and managed by public-relations firm Ketchum of New York, to be a “straight-up marketing tool to spin GMOs in a positive light”. It is now seeking the records of public-sector researchers — who are subject to state freedom-of-information laws — to confirm its suspicions.

Ruskin says that the group has received responses to about 10% of its records requests. At least one institution, the University of Nebraska, has refused to provide documents requested by the group.

US Right to Know argues that its requests are reasonable, because the researchers who are under scrutiny are public employees. “Part of democracy is that we get to know what our public employees do,” says Ruskin.

McGuire is not sure why the group is seeking her records, because she has not contributed to the GMO Answers website. Some ►

1. Bumpa, J. B., Clemens, M. A., Demombynes, G. & Haddadd, L. *Lancet* **379**, 1945 (2012).
2. Pronyk, P. M. *et al. Lancet* **379**, 2179–2188 (2012).
3. Banerjee, A. *et al. Science* <http://dx.doi.org/10.1126/science.1260799> (2015).

► of her recent research refutes claims that glyphosate, a herbicide often used on GM crops, accumulates in breast milk; the work relied on an assay developed with assistance from Monsanto. Still, says McGuire, “I’m a milk-lactation researcher.”

But Folta’s e-mails show him to be a frequent contributor to GMO Answers. Ketchum employees repeatedly asked him to respond to common questions posed by biotechnology critics. In some cases, they even drafted answers for him. “We want your responses to be authentically yours,” one Ketchum representative wrote in a message on 5 July 2013. “Please feel free to edit or draft all-new responses.”

“They thought they could save me time by providing canned answers,” Folta says of his “extremely annoying” Ketchum contacts. “And I don’t know if I used them, modified them or what, but they stopped doing it at some point.” He adds that the correspondence obtained by US Right to Know reveals only a fraction of his work as a scientist, and taken alone does not paint an accurate picture of his work.

Bruce Chassy, a toxicologist at the University of Illinois at Urbana-Champaign who is the subject of two freedom-of-information requests by US Right to Know, says that his e-mails would reveal a similar portrait of “people trying to defend the science against malicious attacks”.

But Chassy acknowledges the ethical questions raised by close relationships between the biotech industry and the public sector. “Are we working for them, or are they working for us?” he asks. “Probably a little bit of both” — in part because universities and companies often have overlapping research interests. US Right to Know aims to reveal this overlap in full.

Michael Halpern, an expert on scientific integrity at the Union of Concerned Scientists in Washington DC, says that Folta’s case suggests that universities should do more to educate researchers on what constitutes a conflict of interest and what types of financial relationship should be disclosed.

“It behooves scientists to disclose their funding sources so there’s no perception of inappropriate influence,” says Halpern. “But that doesn’t mean all private money is tainted or suspect.” ■



Samples from the Ebola epidemic in West Africa are held by public-health agencies in the region and abroad.

INFECTIOUS DISEASE

Biobank planned for Ebola samples

International public-health officials discuss how to maximize research benefits of a widely dispersed collection.

BY ERIKA CHECK HAYDEN

As West Africa’s Ebola outbreak winds down, an effort is under way to make the best use of the tens of thousands of patient samples collected by public-health agencies fighting the epidemic. On 6–7 August, the World Health Organization (WHO) convened a meeting in Freetown, Sierra Leone, to discuss how to establish a biobank for up to 100,000 samples of blood, semen, urine and breast milk from confirmed and suspected Ebola patients, as well as swabs taken from the bodies of people who died from the virus. Held by health agencies in both West

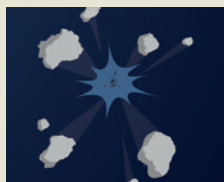
Africa and the West, the samples could be valuable in understanding how the current Ebola crisis evolved, preparing for future outbreaks and developing public-health research capacity in a region that depends on outside experts.

“There are many, many ways that this resource could be precious,” says Cathy Roth, an adviser to the WHO directorate in Geneva, Switzerland, which arranged the meeting as part of a series of international discussions about the creation of an Ebola biobank. One of the difficulties is that there is no blueprint for how such a biobank would work, so countries have not yet committed to joining it.

DANIEL BEREHULAK/NTY/REDUX/EVYME

 **MORE ONLINE**

TOP STORY

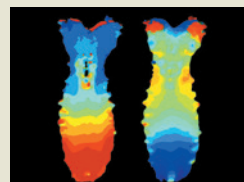


How the rubber-duck comet got its shape
go.nature.com/1u6tgn

MORE NEWS


- 3D-printed device helps computers solve cocktail-party problem go.nature.com/wlrds
- Spotty sunspot record gets a makeover go.nature.com/k53iet
- Chinese biologists blocked after winter Olympics outcry go.nature.com/tdldkj

VIDEO



Fruit-fly brains filmed in action
go.nature.com/riw7x3

KRISTIN BRANSON/WILLIAM LEMON/PHILIPP KELLER/HIMI JANELIA RESEARCH CAMPUS



One proposal has been to link existing collections in an online biobank with a reference laboratory in Africa that would hold certain samples — for instance, collections taken from notable groups of patients, or from people who were followed especially closely throughout the course of their disease. Such a facility would be a first for the region; there is currently no high-containment lab in the Ebola zone that is suitable for studies of live, highly dangerous viruses.

Although the samples vastly outnumber those collected in previous outbreaks, they are still a finite resource. Ongoing discussions will need to grapple with who decides what the samples can be used for and what kinds of research should be emphasized.

“We want to have defined research priorities, because these samples do represent an exhaustible resource,” says Ethan Guillen, project manager of the Ebola initiative at Médecins Sans Frontières (MSF) in Geneva, which is advocating for the biobanking project.

Assigning control over these decisions to the three countries where the majority of Ebola cases occurred — Guinea, Liberia and Sierra Leone — is a major priority for MSF. Historically, much research on viral haemorrhagic fevers such as Ebola has been done by scientists in developed countries using samples taken from developing countries. Guillen sees

that as part of the reason why, 40 years after Ebola was first documented in Africa, there is not enough public-health capacity in some countries to contain the disease or effective tools to treat or prevent it.

Guillen says that affected countries need a system in which “their scientists have a say in what happens and can learn from the experiences, so there doesn’t have to be such a reliance on outside actors”.

Already, thousands of samples have been shipped out of Africa by foreign govern-

“We want to have defined research priorities, because these samples do represent an exhaustible resource.”

ment agencies that stepped in to test patients for Ebola during the outbreak. Several, including the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, the European Mobile

Laboratory Project and the Pasteur Institute in Paris have expressed cautious support for the biobanking idea.

“CDC is supportive of the concept of Ebola biobanks for samples, particularly in the affected countries, which would offer organizations around the world access to samples from the largest Ebola outbreak in history,” the agency told *Nature*.

Only the European mobile lab provided detailed information on how many samples it has — about 3,000 in a high-containment lab in Hamburg, Germany, says virologist Stephan Günther of the Bernhard Nocht Institute for Tropical Medicine in Hamburg, which implements the lab project. Günther says that the European mobile lab is acting as custodian of the samples, which are still owned by the countries in which they were collected. The project has signed agreements with Sierra Leone and Guinea that guarantee access for researchers from those countries, he adds.

Public Health Canada says that it has not yet shipped samples out of West Africa but would not reveal where it is holding them, citing “biosafety and biosecurity concerns”.

The CDC says that it is keeping samples both in West Africa and the United States but would not state how many it holds. In December, it shipped 7,000 from Sierra Leone to the United States (see *Nature* [http://doi.org/6jm; 2014](http://doi.org/6jm;2014)). However, the agency has come under fire recently for major lapses in biosecurity (see *Nature* [http://doi.org/6jn; 2015](http://doi.org/6jn;2015)), and raised this among a number of potential hurdles to creating a biobank.

Guillen says that he is hopeful that these issues can be worked through. “We need better research tools,” he says. “Hopefully we can move quickly to get these tools in place.” ■

Age of the NEUTRINO

BY ELIZABETH GIBNEY
GRAPHIC BY NIGEL HAWTIN

As researchers at CERN, Europe's particle-physics laboratory near Geneva, dream of super-high-energy colliders to explore the Higgs boson, their counterparts in other parts of the world are pivoting towards a different subatomic entity: the neutrino. Neutrinos are more abundant than any particle other than photons, yet they interact so weakly with other matter that every second, more than 100 billion stream — mainly unnoticed — through every square centimetre of Earth. Once thought to be massless, they in fact have a minuscule mass and can change type as they travel, a bizarre and entirely unexpected feature that physicists do not fully understand (see 'An unconventional particle'). Indeed, surprisingly little is known about the neutrino. "These are the most ubiquitous matter particles in the Universe that we know of, and probably the most mysterious," says Nigel Lockyer, director of the

Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois. Four unprecedented experiments look poised to change this. Two — one in China and one in India — already have the go-ahead, and plans to erect detectors in Japan and the United States are in the works (see 'Where they will be detected'). Buried underground to prevent interference from other particles, all four are designed to detect many more neutrinos, and to probe the switching process in more detail, than any existing experiment. The results are expected to feed into some of the most fundamental questions in cosmology (see 'Flurry of experiments'). Some of the experiments will make their own neutrinos; all will use any they can capture from the Sun or from supernova explosions. "The age of the neutrino," Lockyer says, "could go on for a very long time."

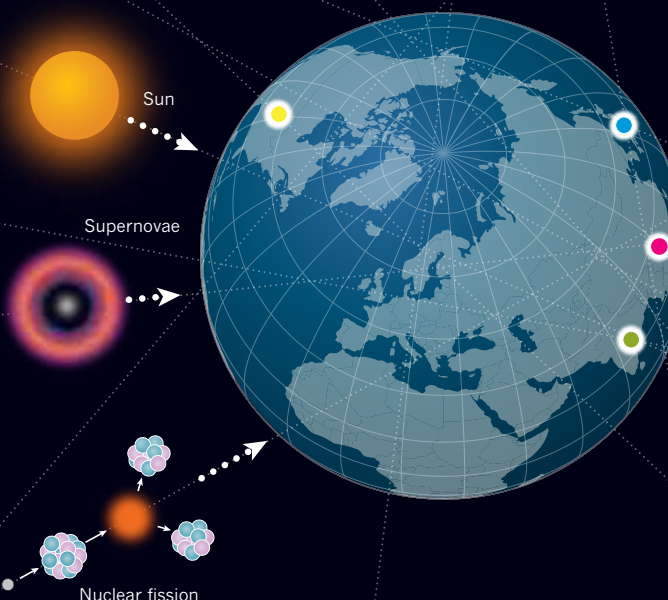
NEUTRINO FACTORIES

Neutrinos are everywhere, generated by a variety of processes.

Fusion of hydrogen nuclei to form helium in the Sun.

Supernovae and collisions between cosmic rays and air particles in Earth's atmosphere.

Particle accelerators smashing protons into a target and fission from the radioactive decay of elements inside nuclear reactors.



WHERE THEY WILL BE DETECTED

Deep Underground Neutrino Experiment (DUNE), United States
Status: Planned
Cost: US\$1 billion
Will make highest-energy neutrinos of any experiment.

Hyper-Kamiokande, Japan
Status: Planned
Cost: About \$800 million
Will be the world's largest neutrino detector — it is 25 times bigger than its predecessor, Super-Kamiokande.

Jiangmen Underground Neutrino Observatory (JUNO), China
Status: Construction begun
Cost: \$330 million
Sits under 700 metres of rock.

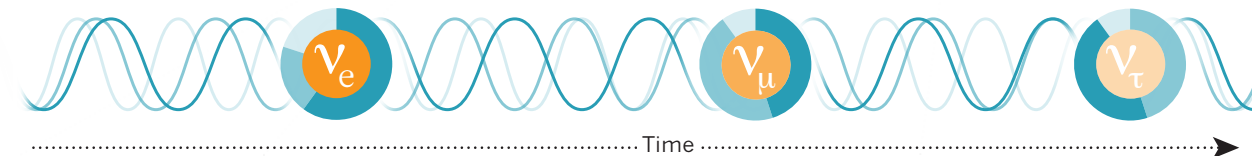
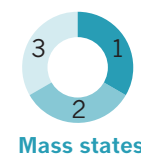
India-based Neutrino Observatory (INO), India
Status: Funding approved
Cost: \$233 million
Will be largest experimental basic-science facility in India.

AN UNCONVENTIONAL PARTICLE

A neutrino (ν), or its antimatter counterpart the antineutrino, is always produced alongside an electron (e) or one of the electron's heavier cousins, the muon (μ) or tau (τ) particle — and the presence of this partner particle gives the neutrino a 'flavour'.



Unlike electrons, muons and tau particles, neutrinos do not have definite masses. Instead, every neutrino is a mixture — or quantum superposition — of three 'mass states', and those states mix in different proportions to make different flavours.



As a neutrino travels, each state contributes to its mass at a varying rate, causing the neutrino to change flavour over time. The frequency of the changes depends on the differences between the mass states, the neutrino's energy and parameters that govern how the states are allowed to mix.

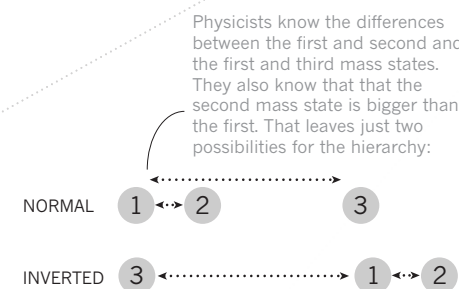
Flurry of experiments

The detectors in China (JUNO) and India (INO) are designed to untangle the relationship between the three mass states, with implications for the origins of the forces of nature. By contrast, DUNE in the United States and Hyper-Kamiokande in Japan aim to spot differences in how neutrinos and antineutrinos oscillate between flavours. That could solve a second cosmological puzzle: why the Universe is made up of matter rather than antimatter. All four detectors will also hunt for a hypothesized 'sterile' neutrino.

BIG QUESTIONS

What is the mass hierarchy?

Although physicists know that neutrinos exist in three different mass states, which state is the lightest and which is the heaviest remains a mystery. Knowing that would help scientists to decide between rival theories about how the four forces of nature unite as a single force at high energies, similar to those experienced in the moments after the Big Bang.



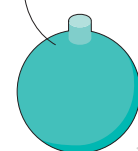
Physicists know the differences between the first and second and the first and third mass states. They also know that the second mass state is bigger than the first. That leaves just two possibilities for the hierarchy:

2020

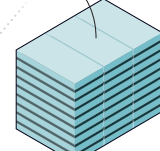
JUNO

Will measure the rate at which antineutrinos of different energies created at the Yangjiang and Taishan nuclear power plants (53 kilometres apart) switch flavour to calculate the differences between mass states.

20,000 tonnes of 'liquid scintillator' lights up when neutrinos hit



50,000 tonnes of magnetic iron plates distinguish neutrino from antineutrino strikes



INO

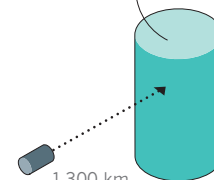
Will detect neutrinos and antineutrinos produced by cosmic rays from the other side of Earth. If the journey boosts neutrino switching, this implies a normal mass hierarchy; if antineutrino switching speeds up, the inverted hierarchy is likely.

2025

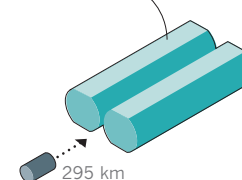
DUNE

Will send neutrinos of different energies from Fermilab to the Sanford Underground Research Facility in South Dakota. Physicists will record differences in the way neutrinos and antineutrinos oscillate and how this depends on their energy.

40,000 tonnes of liquid argon produces electrons and light when neutrinos hit



1 megatonne of water shows cones of light where neutrinos hit



Hyper-Kamiokande

Neutrinos and antineutrinos will travel from the Japan Proton Accelerator Research Complex (J-Parc) in Tokaimura. Particles will be of a single energy, selected to maximize the detection of flavour switching over the distance from J-Parc.



A new generation of economists is trying to transform global development policy through the power of randomized controlled trials.

REVOLT OF THE RANDOMISTAS

BY JEFF TOLLEFSON

In 70 local health clinics run by the Indian state of Haryana, the parents of a child who starts the standard series of vaccinations can walk away with a free kilogram of sugar. And if the parents make sure that the child finishes the injections, they also get to take home a free litre of cooking oil.

These simple gifts are part of massive trial testing whether rewards can boost the stubbornly low immunization rates for poor children in the region. Following the model of the randomized controlled trials (RCTs) that are commonly used to test the effectiveness of drugs, scientists randomly assigned clinics in the seven districts with the lowest immunization rates to either give the gifts or not. Initial results are expected next year. But

smaller-scale experiments suggest that the incentives have a good chance of working. In a pilot study conducted in India and published in 2010, the establishment of monthly medical camps saw vaccination rates triple, and adding on incentives that offered families a kilogram of lentils and a set of plates increased completion rates by more than sixfold¹.

"We have learned something about why immunization rates are low," says Esther Duflo, an economist at the Massachusetts Institute of Technology (MIT) in Cambridge, who was involved in the 2010 experiment and is working with Haryana on its latest venture. The problem is not necessarily that people are opposed to immunization, she says. It is that certain obstacles, such as lack of time or money, are making it difficult for them to attend the clinics. "And you can balance that difficulty with a little incentive," she says.

ESTHER DUFLO/MIT; SABIN VACCINE INSTITUTE



This is one of a flood of insights from researchers who are revolutionizing the field of economics with experiments designed to rigorously test how well social programmes work. Their targets range from education programmes to the prevention of traffic accidents. Their preferred method is the randomized trial. And so they have come to be known as the ‘randomistas’.

The randomistas have been particularly welcomed in the global development arena. Despite some US\$16 trillion in aid having flowed to the developing world since the Second World War, there are little empirical data on whether that money improves the recipients’ lives (see page 144). The randomistas see their experiments as a way to generate such data and to give governments tools to promote development, relieve poverty and focus money on things that work.

Trials are showing that offering incentives can boost attendance at vaccination clinics.

Not everyone is convinced. Sceptics argue that the randomistas’ focus on evaluating specific aid programmes can lead them to lose sight of things such as energy, infrastructure, trade and corruption — macroeconomic issues that are central to a country’s ability to prosper, but that are effectively impossible to randomize. “Development is ultimately about politics,” says Angus Deaton, an economist at Princeton University in New Jersey.

Nonetheless, the randomista movement is gaining momentum (see ‘Scale the heights’). Universities are pumping out more economics graduate students with experience in RCTs every year. Organizations ranging from the UK Department for International Development to the Bill & Melinda Gates Foundation in Seattle, Washington, are throwing their financial support behind the technique. “There are hundreds and hundreds of randomized trials going on, and ten years ago that just wasn’t the case,” says economist Dean Karlan at Yale University in New Haven, Connecticut, who is at the forefront of the movement. “We’ve changed the conversation.”

Demand is only rising. This September, governments will gather in New York under the auspices of the United Nations to approve a new set of Sustainable Development Goals, which are intended to guide investments over the coming decade. And in December, questions about financial aid will be high on the agenda at the UN climate summit in Paris, where governments expect to sign a new climate agreement that will probably include commitments by industrialized nations to funnel money into sustainable development in poorer countries. In both cases, the effectiveness of the programmes is likely to be a key concern.

“This is front and centre on a lot of people’s agenda,” says Ann Mei Chang, who is executive director of the Global Development Lab at the US Agency for International Development (USAID) in Washington DC. “Where do we get the biggest bang for our buck?”

PROGRESS AND OPPORTUNITIES

RCTs have been used to test the effectiveness of social programmes at least since the 1960s. But the modern era began in 1997, when one of the most famous and influential RCTs in public policy began in Mexico.

The experiment had its origins three years earlier, when Mexican President Ernesto Zedillo assumed office in the middle of an economic crisis and assigned economist Santiago Levy to devise a programme to help poor people. Sceptical of the conventional approach — subsidies for products such as tortillas and energy — Levy designed a system that would provide cash payments to poor families if they met certain requirements, such as visiting health clinics and keeping their children in school. “And because people were very critical about what I was doing,” says Levy, who now leads strategic development planning at the Inter-American Development Bank in Washington DC, “I wanted to ensure that we had numbers so that we could have an informed debate.”

As it happened, Levy had a natural control group for his experiment. The government was rolling out its payment programme in stages, so he could collect data on families in villages that were included in the initial roll-out, and in comparable villages that were not. Within a few years, his team had data suggesting that the programme, dubbed *PROGRESA*, was working remarkably well. Visitation to health clinics was 60% higher in participating communities than in the control group. Children in those communities also had a 23% reduction in illness and an 18% reduction in anaemia. Overnight hospital visits halved across several age ranges.

These data helped to solidify support for the programme. Now known as *Prospera*, it covers almost all of Mexico’s poorest citizens and has inspired similar initiatives across Latin America and into Africa.

“*PROGRESA* was one of the first major national programmes of its kind to get a rigorous evaluation,” says William Savedoff, who works on aid effectiveness and health policy at the Center for Global Development, a think tank in Washington DC. “Today conditional cash-transfer programmes are some of the most heavily evaluated programmes in the world, and that is I think a direct consequence of the Mexican experience.”

The idea of developing hard evidence to test public policies was

bubbling up in parallel in the United States. One of the first trials began in 1994 with a small initiative to analyse the effect of supplying textbooks and uniforms as well as basic classroom improvements to a group of schools in Kenya. Economist Michael Kremer at Harvard University in Cambridge had taught in Kenya years earlier. A friend of his who worked for a non-profit group was initiating the programme, and Kremer suggested that the group roll it out as an experiment. “I didn’t necessarily expect anything to come of this,” he says.

Working with the group, Kremer collected data on students in 14 schools, half of which received the intervention. School attendance increased, but test scores did not. Similar results came from an experiment in 1995 that involved 100 schools. That trial suggested that providing textbooks had little effect on average test scores², owing perhaps to language challenges — the textbooks were in English, which was not the native language for many students. Students who were already scoring higher than their peers, however, pulled further ahead if they had the books.

Kremer continued to run RCTs of other programmes, but it was Duflo — then a student of his — who pushed the idea into the mainstream. Duflo’s 1999 dissertation looked in part at an education initiative in Indonesia that had built 61,000 primary schools over 6 years in the 1970s. She wanted to test a common concern that such a rapid expansion would lead to a decline in the quality of education, thereby offsetting any gains. Running an experiment was impossible, but Duflo was able to use data on the differences across regions to show that the programme had, in fact, increased educational opportunities as well as wages.

This and other early work inspired Duflo to look at RCTs as a way to generate data and definitively measure the effectiveness of policies and programmes. “As soon as I had a longer time horizon and some money I started working on setting some up,” she says.

One of Duflo’s early papers³, published in 2004, capitalized on a 1993 amendment to India’s constitution that devolved more power over public investments to local councils and reserved the leadership of one-third of those councils, to be chosen at random, for women. Duflo realized that this effectively created a RCT that could test the effect of having women-led councils. In analysing the data, she found that councils led by women boosted political engagement by other women and directed investment towards issues raised by them. In some areas, women are in charge of obtaining drinking water, for instance, and councils led by women typically invested more in water infrastructure than did those run by men. “The scale of the policy and the topic were at the time unusual,” Duflo says. “It gave me a sense of the range of things that the tool could possibly cover.”

By the early 2000s, the randomistas were on the upswing. In 2002, Karlan, one of Duflo’s students, joined with her and other researchers to form Development Innovations — now known as Innovations for Poverty Action — in New Haven. The following year, Duflo co-founded what is now known as the Abdul Latif Jameel Poverty Action Lab (J-PAL) in Cambridge with fellow MIT economists Abhijit Banerjee and Sendhil Mullainathan.

The work quickly expanded, and J-PAL has now run nearly 600 evaluations in 62 countries, and trained more than 6,600 people. One of Duflo’s latest projects will revisit her dissertation on education in Indonesia, only this time with secondary schools and randomized control groups. “We will have a randomized version of a paper on the benefits to education soon I hope,” Duflo says.

VENTURE CAPITAL

One enthusiastic convert to the randomista philosophy is Rajiv Shah, a Gates Foundation official who became head of USAID in 2010. Once there he created a fund called Development Innovation Ventures (DIV) to test and scale up solutions to development problems, and he enlisted Kremer as its scientific director. The goal, Shah said, was to

“move development into a new realm” through the use of evidence.

Since then DIV has invested in more than 100 development projects, and nearly half involve RCTs. One, conducted in Kenya by a pair of researchers from Georgetown University in Washington DC, tested a simple method for reducing traffic accidents that involve minibuses — collisions that Kremer calls major and increasing killers. “Two of them crash into each other, and 40 people die,” he says.

In 2008, the researchers worked with more than 1,000 drivers to place stickers on buses that urged passengers to speak up about reckless driving⁴. They then collected information from four major insurance companies and found that claims for serious accidents had dropped by 50% on buses with stickers compared with those without. DIV provided a grant to conduct a larger trial — which found that claims dropped by 25–33% — and a second grant of nearly \$3 million to help to scale up the project throughout Kenya.

“The really big win is when developing countries, or firms or NGOs [non-governmental organizations] change their policies,” Kremer says. But one question now facing DIV is whether such a strategy — or indeed any project that proves effective in one setting — can be repackaged and deployed in other countries, where different cultural factors are at play (see *Nature* 523, 516–518; 2015).

SCALE UP

Effecting policy change is the precise aim of the Global Innovation Fund, which was launched in September 2014 with \$200 million over 5 years from the UK Department for International Development, USAID and others, and which follows the DIV model of rigorous testing. Interim director Jeffrey Brown, who is on loan from USAID, says that the fund has already received more than 1,800 applications for projects in 110 different countries and will be announcing its first suite of grants later this year. “We are essentially trying to become a bridge over the valley of death for good development ideas,” he says.

But such organizations still provide only a tiny fraction of the billions of dollars that are spent each year on development aid, let alone the trillions of dollars that are spent by governments on domestic social programmes. Even at lending institutions that have taken this evidence-based framework on board, the portion of investments that is covered by rigorous evaluations is small.

At the World Bank, which started a Development Impact Evaluation division in 2005, the number of projects receiving formal impact evaluations — through RCTs or other means — rose from fewer than 20 in 2003 to 193 in 2014, mostly covering things such as agriculture, health and education. But that still represents just 15% of the bank’s projects, says evaluation-division head Arianna Legovini, who leads a team of 23 full-time staff and has an annual budget of roughly \$18 million. Although many of these evaluations more than pay for themselves over the long term, one constraint is the up-front cost: the average price of an impact evaluation is around \$500,000. “If I did not have donor funding,” she says, “these studies just would not happen.”

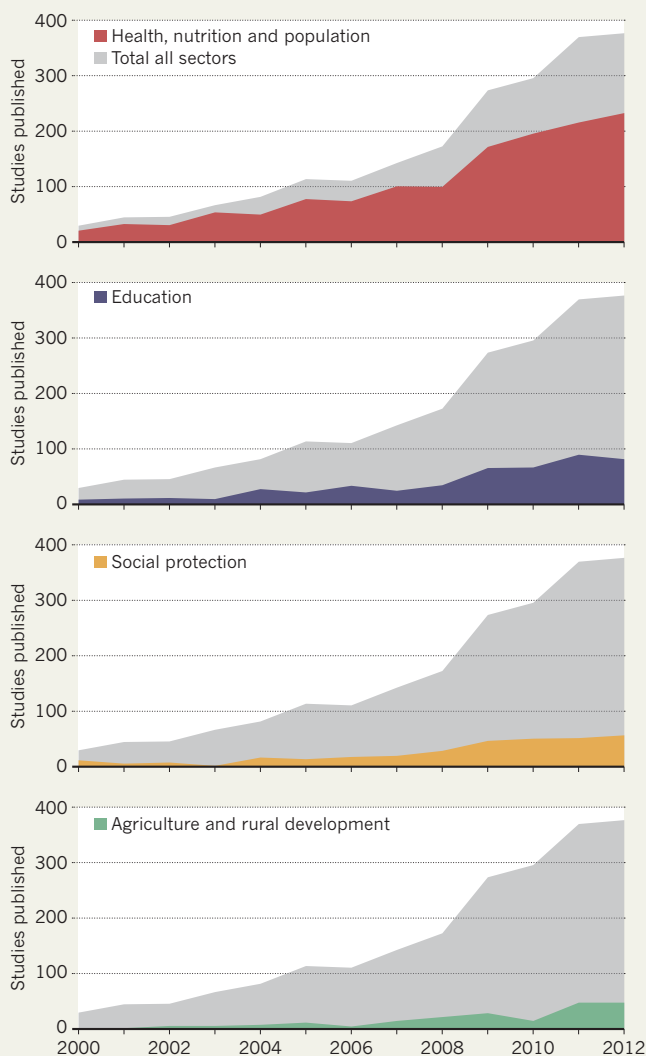
The World Bank is trying to make the most of its resources by working directly with developing countries on implementation. More than 3,000 people have attended its workshops and training sessions since 2005, most of whom were government officials in developing countries that are receiving funds from the bank.

The bank is also making efforts to assess the impact-evaluation programme itself — although the analysis is based largely on whether payments for projects are made on time as a proxy for implementation of the initiatives. An analysis by Legovini and two of her team suggests that development projects that undergo a formal impact analysis are more likely to be implemented on time than are those that do not have evaluations, probably because of the extra attention

**“THE FAD NOW IS LET’S
PILOT IT, AND IF IT WORKS
WE’LL TAKE IT TO SCALE.”**

SCALE THE HEIGHTS

The growing influence of the randomized controlled trial in economic spheres can be seen in the number of studies published each year. Most of the increase is in four sectors — although many studies overlap.



D. B. CAMERON, A. MISHRA AND A. N. BROWN, J. DEV. EFF. HTTP://DOI.ORG/GN8 (2015).

that is given to initial set-up, roll-out and monitoring⁵.

This finding is good news for individual projects, but it is also a potential thorn in the side of many RCTs. Positive effects seen in a trial setting may disappear when the programme is scaled up, governments take over and all the extra attention disappears (see *Nature* **523**, 146–148; 2015).

“The fad now is let’s pilot it, and if it works we’ll take it to scale,” says Annette Brown, who heads the Washington DC office of the International Initiative for Impact Evaluation, an organization that funds impact evaluations as well as meta-analyses of existing studies. Brown says that researchers and governments should probably conduct rigorous studies when any programme is scaled up to ensure that the results continue to hold true — just as the government in Haryana is doing now.

RANDOMIZATION BIAS

From a political perspective, the strongest argument in favour of well constructed RCTs — that they do not lie — may also be the biggest factor working against them. Local politicians often want to cut ribbons and release money into communities, whereas international donors, including governments and NGOs, want flagship programmes that show how they are improving the world. They do

not welcome results showing that initiatives are not working. Even in Mexico, Levy says, some of the subsidies that he fought against when he created *PROGRESA* have regained political favour.

But the randomistas have been accused of succumbing to their own biases. Some fear that their insistence on the RCT has skewed research towards smaller policy questions and given short-shrift to larger, macroeconomic questions. One example comes from Martin Ravallion. An economist at Georgetown University and a former research director at the World Bank, he cites an antipoverty programme in China that received \$464 million from the bank in the 1990s. Although the programme involved road construction, housing, education, health and even conditional cash payments for poor families, a study based on data collected in 2005, 4 years after disbursement ended, found minimal average impact on citizens⁶. “That was the only long-term study of integrated rural development, which is the most common form of development assistance,” Ravallion says.

Yet some families did benefit, and by combining statistics with economic modelling, he and his team showed that the difference lay in basic issues, such as education level. For Ravallion, the message is that aid is best targeted at the literate poor, or more broadly at issues such as literacy. “Governments need to know these things,” he says. “They can’t just know about the subset of things that are amenable to randomization.”

To Alexis Diamond, a former student of Duflo’s who manages project evaluations at the International Finance Corporation, the private-sector development arm of the World Bank in Washington DC, the debate between the randomistas and the old-guard economists is in many ways about status and clout. The latter have spent their careers delving into ever more complex and abstract models, he says. And then “the randomistas came along and said ‘We don’t care about any of that. This is about who has a seat at the table.’”

Diamond says that he tries to strike a balance at his organization, where most evaluations still rely on a mixture of quantitative and qualitative data, including expert judgement.

Duflo shrugs off the debate and says that she is merely trying to provide government officials with the information — and tools — that they need to help them spend their money more wisely. “The best use of international aid money should be to generate evidence and lessons for national governments,” she says.

She points to an anti-pollution programme in industrial plants in the Indian state of Gujarat. Partnering with a group of US researchers, the state ran an experiment in 2009 that divided nearly 500 plants into 2 groups. Those in the control group continued with the conventional system, in which industries hire their own auditors to check compliance with pollution regulations. The others tested a scheme in which independent auditors were paid a fixed price from a common pool. The hope was that this would eliminate auditors’ fear of being black-balled for filing honest reports. And it did: independent auditors were 80% less likely to falsely give plants a passing grade, and many of the industrial plants covered by those audits responded by curbing their pollution. In January, regulators rolled out the programme across the state.

“My hope, in a best-case scenario, is that in the next ten years you are going to have many, many of these projects run as a matter of course by governments in the spaces where they want to learn,” Duflo says. ■ [SEE EDITORIAL P.135](#)

Jeff Tollefson is a reporter for *Nature* in New York.

1. Bannerjee, A. V., Duflo, E., Glennerster, R. & Kothari, D. *Br. Med. J.* **340**, c2220 (2010).
2. Glewwe, P., Kremer, M. & Moulin, S. *Am. Econ. J. Appl. Econ.* **1**, 112–135 (2009).
3. Chattopadhyay, R. & Duflo, E. *Econometrica* **72**, 1409–1443 (2004).
4. Habyarimana, J. & Jack, W. *Heckle and Chide: Results of a Randomized Road Safety Intervention in Kenya* (Center for Global Development, 2009); available at go.nature.com/lyewbv
5. Legovini, A., Di Maro, V. & Piza, C. *Impact Evaluation Helps Deliver Development Projects* (World Bank Group, 2015); available at go.nature.com/tv3hqq
6. Chen, S., Mu, R. & Ravallion, M. *J. Publ. Econ.* **93**, 512–528 (2009).

COMMENT

EMBODIED COGNITION How our bodies influence our thinking **p.158**



PARTICLE PHYSICS Mary K. Gaillard's convoluted track to tenure **p.160**

ENERGY Plant food crops alongside panels in solar farms **p.161**

EDUCATION Support teaching academics and retain factual knowledge in classes **p.161**

JIANG KEHONG/XINHUA PRESS/CORBIS



Discarded shrimp shells contain nutrients that could be used to enrich animal feed.

Don't waste seafood waste

Turning cast-off shells into nitrogen-rich chemicals would benefit economies and the environment, say **Ning Yan** and **Xi Chen**.

Every year, some 6 million to 8 million tonnes of waste crab, shrimp and lobster shells are produced globally — about 1.5 million tonnes in south-east Asia alone¹. Whereas 75% of the weight of a tuna fish can be extracted as fillets, meat accounts for only around 40% of a crab's mass.

In developing countries, waste shells are often just dumped in landfill or the sea. In developed countries, disposal can

be costly — up to US\$150 per tonne in Australia, for example.

Yet shells harbour useful chemicals — protein, calcium carbonate and chitin, a polymer similar to cellulose, but which contains nitrogen (see 'Shell biorefinery'). The potential value of such shells for the chemical industry is being ignored. Scientists should work out sustainable ways to refine crustacean shells, and governments and industry should invest in using this

abundant and cheap renewable resource.

Dried shrimp shells are valued at a mere \$100–120 per tonne. They can be ground down and the powder used as an animal-feed supplement, bait or fertilizer, as well as in chitin production. The return is not much more for agricultural residues and wastes: corn stover and wheat straws, which are burned for heat or refined into chemicals, sell for \$50–90 per tonne.

Crustacean shells are 20–40% ▶

► protein, 20–50% calcium carbonate and 15–40% chitin. What could these parts be used for?

Protein is good for animal feeds. For example, *Penaeus* shrimp shells contain all the essential amino acids and have a nutrient value comparable to that of soya-bean meal. Today, the protein is not being used because the current processing methods destroy it. As livestock breeding rises rapidly, waste crustacean shells from southeast Asia could be transformed into protein-rich animal feed with an annual market value of more than \$100 million, according to World Bank data.

Calcium carbonate has extensive applications in the pharmaceutical, agricultural, construction and paper industries. It currently comes mainly from geological sources such as marble and limestone. These sources are plentiful but might contain heavy metals that are difficult to remove. Chalk from shells would thus be better for human consumption, for example as a constituent of pills. People might also find it easier to accept tablets that originate from food sources than from rocks.

The market price of ground calcium carbonate is around \$60–66 per tonne for coarse particles, which are used in construction, pigments, fillers and soil treatments. Ultrafine particles, which can be used to improve the properties of rubber and plastics, can reach an astonishing \$14,000 per tonne. Even if the calcium carbonate from southeast Asian crustacean shells was processed into only the cheapest coarse particles, it could have an annual market value of up to \$45 million.

Chitin is a linear polymer and the second most abundant natural biopolymer on Earth (after cellulose). It is found in fungi,

plankton and the exoskeletons of insects and crustaceans, and organisms generate about 100 billion tonnes of chitin every year². Currently, the polymer and its water soluble derivative, chitosan, are used in only a few niche areas of industrial chemistry, such as cosmetics, textiles, water treatment and biomedicine. Its potential is much greater.

Unlike most other forms of biomass such as cellulose, chitin contains nitrogen. Nitrogen-containing compounds — widely used in the pharmaceutical industry, carbon dioxide fixation, textiles and beyond — are crucial for modern life. For example, the nitrogen-containing organic compound pyrazine is integral to several best-selling drugs such as eszopiclone (for sleeping difficulties) and varenicline (to treat nicotine addiction). Ethanolamine (ETA) is used in power plants for CO₂ sequestration and in skin-friendly soaps, household cleansers and surfactants. Nitrogen-containing chemicals have a huge market — about 2 million tonnes of ETA are used a year globally³, with annual sales of around \$3.5 billion.

The industrial production of nitrogen compounds involves fossil fuels and energy-intensive processes. First, nitrogen gas must be converted into ammonia through the Haber process, which is notorious for its low reaction efficiency. This process alone accounts for an estimated 2–3% of global energy consumption. For every mole of nitrogen gas consumed,

three moles of hydrogen gas, derived from fossil fuels, are used.

Further processing is complex and expensive. For instance, generating ETA requires six steps: hydrogen production from coal or natural gas; nitrogen isolation from air; ammonia synthesis; ethylene production from crude-oil cracking; conversion of ethylene into ethylene oxide; and then conversion of ethylene oxide into ETA.

Chitin might be a more suitable starting point for ETA production. With carbon, nitrogen and oxygen already bound in the polymer, only one step is needed to make ETA. Another five chemicals have been derived from chitin in a single step and the list is growing. So far, however, this has been achieved on only a small scale in the lab⁴.

CHEMICAL CHALLENGES

Extracting chemicals from waste shells with existing methods is destructive, wasteful and expensive. It requires separating out the different components, a process known as fractionation. Protein is removed with sodium hydroxide solution and the decomposition of calcium carbonate uses hydrochloric acid — both are corrosive and hazardous solvents.

To make chitosan, chitin is treated with 40% concentrated sodium hydroxide solution. The production of 1 kilogram of chitosan from shrimp shells requires more than 1 tonne of water.

As a result, good quality chitin can cost up to \$200 per kilogram, although the starting material is cheap. The global industrial use of refined chitin (in membranes, drug delivery, food and cosmetics) is low: around 10,000 tonnes per year³. Few chitin facilities exist; China, Japan, Thailand and Indonesia have a few. The transformation of chitin or chitosan to other chemicals poses further problems. Natural chitin is a crystalline material that prevents reagents from easily accessing the polymer chains. Under harsh reaction conditions, the chains easily undergo side reactions to form myriad complex compounds. Separation of the bio-based products from the reactor is often laborious.

In our view, these challenges are no greater than those in processing woody biomass into biofuels and other chemicals, which took two decades to move from the lab to commercial scales.

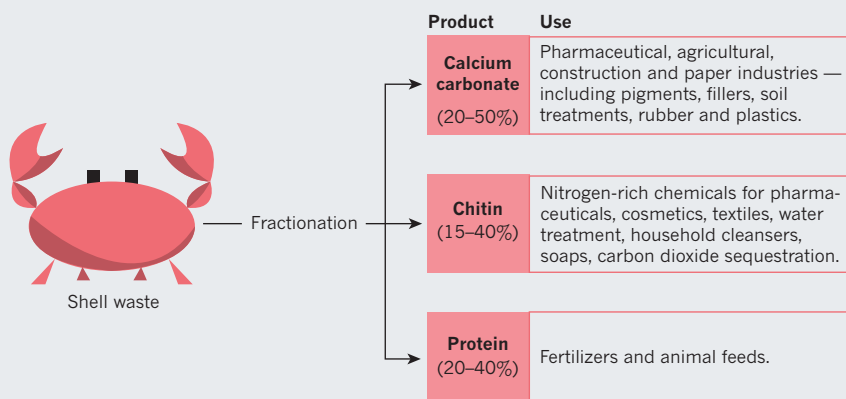
Establishing a profitable, sustainable industry from shell waste is going to take creative chemistry. It needs a sustainable fractionation method to separate proteins, calcium carbonate and chitin — one that avoids corrosive or hazardous reagents and minimizes waste.

New technologies are emerging. For example, teams in Mexico and the United Kingdom demonstrated a lactic-acid

“In the next five years, a multimillion-dollar project should be launched to establish the first processing pipeline.”

SHELL BIOREFINERY

Crustacean shells contain three primary chemicals that have many industrial uses. Developing a sustainable way to refine them could add billions of dollars to the bioeconomy.



fermentation process for chitin production in the lab and in a pilot plant in the early 2000s^{5,6}. The process converted up to 30–50 kilograms of shell waste in a single reactor. A mixture of bacteria that consumes proteins and decomposes calcium carbonate has been developed by groups in the United Kingdom, United States and China^{6–8}. Protein hydrolytes and calcium lactate are by-products that are useful for animal feed and calcium supplements.

GET CRACKING

Another option would be to design and use ionic liquids (liquid organic compounds with ionic functional groups) that can dissolve carbohydrate polymers and extract chitin. Chitin polymers produced in this way have long chains and a high molecular weight, and can be spun into fibres and films for wound dressings and water treatment, for example.

Researchers also need to explore physical, solvent-free methods for shell fractionation. Ball milling (placing materials with metal balls in a spinning cylinder) may be used to grind the shells finely and break apart crystals. Combining chemical and mechanical forces might prove advantageous. For example, using a ball mill and an acid catalyst can degrade wood without heating. Combining steam explosion (a technique that uses superheated steam and sudden pressure release) with acid is another way to liberate the shell's components.

Ball milling and steam explosion have been used for woody biomass refining at a pilot scale but few people have noticed the potential of these techniques for waste shells. (In collaboration with the Chinese Academy of Sciences' Institute of Process Engineering in Beijing, our group at the National University of Singapore aims to have a pilot demonstration for shells running in a few years.)

Converting chitin into small nitrogen-containing chemicals — such as derivatives of ETA and of the widely used organic solvent furan⁹ — is developing fast, although is still in early stages. It may take at least five years to scale up the process and another ten years to commercialize it. Future investigations need to explore routes from chitin to other chemicals, enhance product yields through improved catalysis and pre-treatments and ease the separation of products.

We propose that a processing pipeline be developed for refining waste shells, just as woody biomass (composed mainly of cellulose, hemicellulose and lignin) is separated and converted into a range of products in one facility¹⁰. That development took the cooperation of many parties, propelled by public concern over energy security and climate change. It also required financial



Lobster shells harbour the nitrogen-rich compound chitin, which is used in pharmaceuticals.

support from governments and the chemical and fuel industry. Shell-waste biorefineries will create new industrial opportunities in southeast Asia and beyond.

Strong support from policymakers, research institutes, governments, funders and the public is key. Fundamental research from scientists worldwide is urgently needed to overcome the technical barriers.

SHELL REFINERY

In the next five years, a multimillion-dollar project should be launched to establish the first processing pipeline using new technology. The project should be supported by governments of nations rich in shell waste, and executed by researchers with complementary expertise, covering catalysis, materials science and engineering, food science and life-cycle assessment. Companies — including producers and traders of shellfish, those associated with biocommodities and biomaterials and others promoting renewable materials — should reassess the potential markets of an environmentally friendly and profitable waste-shell refinery and engage with research to commercialize emerging technologies.

In the next decade, stringent regulations

should be implemented on the disposal of waste shells, while providing incentives for companies who use them. ■

Ning Yan is professor of green chemistry and **Xi Chen** is a research fellow in the Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore.
e-mail: ning.yan@nus.edu.sg

1. Food and Agriculture Organization of the United Nations. *The State of World Fisheries and Aquaculture* (FAO, 2014).
2. Kim, S. K. *Chitin, Chitosan, Oligosaccharides and Their Derivatives: Biological Activities and Applications* (CRD Press, 2011).
3. Grand View Research. *Amines Market Analysis By Product (Ethanolamines, Alkylamines, Fatty Amines), By Application (Crop Protection, Surfactants, Water Treatment, Personal Care) And Segment Forecasts To 2020* (Grand View Research, 2014); available at go.nature.com/ofjys4
4. Chen, X. & Yan, N. *Catal. Surv. Asia* **18**, 164–176 (2014).
5. Cira, L. A., Huerta, S., Hall, G. M. & Shirai, K. *Process Biochem.* **37**, 1359–1366 (2002).
6. Beaney, P., Lizardi-Mendoza, J. & Healy, M. *J. Chem. Technol. Biotechnol.* **80**, 145–150 (2005).
7. Zhang, H. et al. *Carbohydr. Res.* **362**, 13–20 (2012).
8. Liu, P. et al. *Biochem. Eng. J.* **91**, 10–15 (2014).
9. Bobbink, F. D. et al. *Green Chem.* **17**, 1024–1031 (2015).
10. Ragauskas, A. J. et al. *Science* **311**, 484–489 (2006).

A grasp on human thinking

Elsbeth Stern weighs up two studies probing the idea of the brain as the body's servant.



Rock art in the Cave of the Hands, Argentina, dating to between 9,500 and 13,000 years old.

How has *Homo sapiens* uncovered the laws of nature, invented technology and established culture and institutions? Most scientists' answers to that question have been top-heavy, referring to language, symbolic reasoning and consciousness as unique human abilities on which comprehension, analysis, abstraction and reasoning are based. Since the 1950s, those abilities have increasingly become a focal point for psychological research. Encouraged by progress in informatics, researchers began to create digital models of the processes by which sensory input is selected by the brain, stored in the memory, connected to existing knowledge and used for elaboration. These 'cognitive architectures' were supposed to simulate and predict learning, reasoning, complex problem-solving and decision-making.

This algorithmic focus on mental activities ignores the fact that human beings engage with evolutionary pressures using their entire bodies — a point explored by

psychologist Guy Claxton in *Intelligence in the Flesh*, and by philosopher Colin McGinn in *Prehension*.

Intelligence in the Flesh deals with the unity of mind, brain and body in human information-processing, including higher cognition and academic learning. Claxton argues that humans would think and behave differently if their physiological functioning were different. For instance, there is research that shows how holding a cup of hot coffee or receiving other sensory input through the skin can influence judgement and decision-making (L. E. Williams and J. A. Bargh *Science* 322, 6060–607; 2008), a fact entirely ignored in cognitive theories that confine themselves to visual and auditory input. The brain coordinates information, but it is the "servant, not master of the body", notes Claxton.

McGinn's focus in *Prehension* is the human hand. He is not the first to emphasize that thanks to their bipedal gait, early humans did not need their 'forepaws' for

Intelligence in the Flesh: Why Your Mind Needs Your Body Much More Than it Thinks

GUY CLAXTON

Yale Univ. Press: 2015.

Prehension: The Hand and the Emergence of Humanity

COLIN MCGINN

MIT Press: 2015.

locomotion, freeing them to manipulate the environment with the help of tools. However, McGinn goes further, positing that the multiple opportunities provided by our hands shape our concepts of the mind. Therefore we conceive cognitive processes in manual terms, such as 'grasping an idea'.

Claxton and McGinn value higher-order cognition and academic learning differently. McGinn argues that the close interaction between brain and hand allowed humans to find their evolutionary niche through the discovery of physical tools, as well as mental ones such as language or mathematical symbols. He claims that using the hands for pointing and communicating resembles 'air writing', and thereby facilitated the invention of script. Claxton, by contrast, thinks that cognitive competencies based on symbolic systems such as writing (which he pejoratively labels "Cartesian education" in reference to philosopher René Descartes's idea of mind-body dualism) are overvalued, whereas handicrafts and vocational education are undervalued. A strong focus on academic learning and abstract reasoning does not meet the needs of the majority, he argues — to the point that this form of intelligence is essentially alien to humans. Meanwhile, McGinn posits that it is why our otherwise sparsely equipped species has survived.

There is a bullish flavour to their modes of argument that shows that Claxton and McGinn are aware of how controversial their claims are. In fact, how new and robust is the science in each book? Criticism of the shortcomings of cognitive architectures is no novelty. Since the 1980s, the evolutionary aspects of human behaviour and cognition have become a seminal topic throughout psychology. It is widely acknowledged that humans are challenged by the fact that we are adapted to the world as it existed more than 30,000 years ago. It is fully accepted that we are born endowed with perceptual and behavioural programs that were

adaptive for our earliest ancestors and that still affect our behaviour, information-processing and emotional functioning. So when criticizing a Cartesian view of human learning, the authors are preaching to the converted.

Both books are slippery in their dealings with state-of-the-art research. McGinn almost entirely ignores empirical psychology research and instead provides evidence based mostly on plausibility — for instance, when he claims that humans have privileged access to geometry because they can form circles and triangles with their fingers. There is prominent research confirming his emphasis on the pivotal role of hand-brain interaction in human cognition, including dozens of studies on the importance of gesturing in learning. Psychologist Susan Goldin-Meadow's *Hearing Gesture* (Harvard University Press, 2003) is one. McGinn also refers to the views of developmental psychologist Jean Piaget regarding sensorimotor activity as the foundation of cognition in early child development. Yet for more than 30 years, psychologists have shown that the brains of newborns are endowed with core knowledge that prepares them to represent information about objects, quantities and actions long before they can grasp with their hands.

Claxton cherry-picks from psychology and neuroscience literature. When he attacks conventional school education, he provides anecdotal evidence about unhappy children, but ignores evidence-based attempts to improve schooling — for instance, by bringing everyday experience into the teaching of science

“Human beings engage with evolutionary pressures using their entire bodies.”

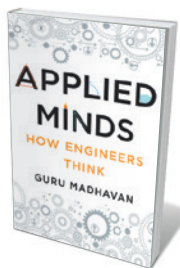
and mathematics. Claxton's claim that performance in intelligence tests is unrelated to factors important in real life is not reflected in

state-of-the-art research, such as the more than 100 publications based on studies of the Lothian Birth Cohort, headed by psychologist Ian Deary of the University of Edinburgh, UK. Intelligence, Deary has shown, is not only significantly related to educational and professional outcome, but is also a factor in positive well-being, health and longevity.

Intelligence in the Flesh and Prehension are eloquently written, refreshing and entertaining. But Claxton and McGinn fight many straw men, and often fail to provide evidence for provocative statements. ■

Elsbeth Stern is a psychologist and professor of teaching and learning research at the Swiss Federal Institute of Technology in Zurich. e-mail: elsbeth.stern@ifv.gess.ethz.ch

Books in brief



Applied Minds: How Engineers Think

Guru Madhavan W. W. NORTON (2015)

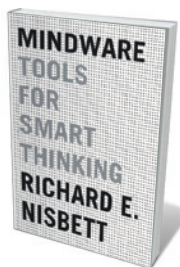
Engineers are titans of real-world problem-solving, yet are strangely invisible, notes biomedical engineer Guru Madhavan. In this riveting study of how they think, he puts behind-the-scenes geniuses such as Margaret Hutchinson, who designed the first penicillin-production plant, centre stage. And, in a feat of reverse engineering, he shows how engineers' methodology — rigorous analysis, testing and orientation towards solutions — is bedded in modular systems thinking, a mindset strong on visualizing structure, designing under constraints and weeding out weak goals in trade-offs.



A River Runs Again: India's Natural World in Crisis, from the Barren Cliffs of Rajasthan to the Farmlands of Karnataka

Meera Subramanian PUBLICAFFAIRS (2015)

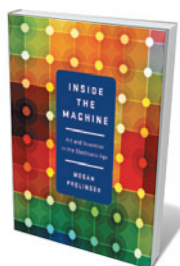
In the middle of India's boom, malnutrition among Indian children is rife. Journalist Meera Subramanian, in search of sustainable solutions for the subcontinent's 1.2 billion people, criss-crossed it to meet scientists and citizens grappling with familiar dilemmas such as child marriage and polluting cooking stoves. Subramanian's analysis is fresher when she takes on the inefficiencies and worse of 'big aid', and her mapping of the micro solutions — such as village rainwater collection — suited to a country of small enterprises.



Mindware: Tools for Smart Thinking

Richard E. Nisbett FARRAR, STRAUS & GIROUX (2015)

How do we decide whether theories are sound or knowledge is only conjectural? Social psychologist Richard Nisbett has drilled into decision-making to produce this “cognitive tool kit” of principles and ideas to aid the process. Inspired by the “seamless web” of science — the interdisciplinary seepage of methods and facts — he draws on economics, psychology and logic for a rich haul. Expect insights in areas ranging from the role of conformism in energy use to the differences in Eastern and Western thinking, and tools from basic statistics to the multipurpose heuristic KISS (keep it simple, stupid).



Inside the Machine: Art and Invention in the Electronic Age

Megan Prelinger W. W. NORTON (2015)

When electronics took off in the 1930s, US technology companies were suddenly forced to convey ‘invisible science’ visually. A bold brigade of commercial artists began to tackle the physics and components with creative brio — but this flowering withered in the 1960s, when the workings of electronics had been absorbed into the culture. For this unusual and compelling study, cultural historian Megan Prelinger has gathered a trove of superb examples. Some are patently influenced by abstract artists such as Wassily Kandinsky, others by surrealism, concrete poetry and science-fiction illustration.



Mess: One Man's Struggle to Clean Up His House and His Act

Barry Yourgrau W. W. NORTON (2015)

Nineteenth-century philosopher Ralph Waldo Emerson wrote, “Things are in the saddle, / And ride mankind”. A thought resonant in a consumerist era, it might also be seen as a comment on hoarding, a condition now associated with obsessive-compulsive disorder. In a memoir mixing sorrow and hilarity, self-confessed clutterer Barry Yourgrau records how he jettisoned junk and traumatic memories by joining Clutterers Anonymous, poking at relevant neuroscience and working his way towards a rapprochement with things. [Barbara Kiser](#)



Mary K. Gaillard's theoretical-physics achievements include predicting the mass of the charm quark.

PHYSICS

She did it all

Val Gibson enjoys the autobiography of Mary K. Gaillard, the first female physics professor at Berkeley.

The brilliant theoretical physicist Mary K. Gaillard has made major contributions to the standard model of particle physics and to superstrings, a candidate theory of everything. In 1981, she became the first woman with a tenured position in the physics faculty at the University of California, Berkeley. Her frank autobiography, *A Singularly Unfeminine Profession*, is an honest, revelatory account of her many discoveries, made as she battled gender bias and faced the demands of raising three children.

Born in New Jersey in 1939, Gaillard has a "survival mechanism" born from an inherent belief in equality, nurtured by her parents and school, and a rebellious tendency to question the world around her. Having fallen in love with physics at school, she won a scholarship to Hollins College near Roanoke, Virginia. It included a year in Paris at l'École Polytechnique — her first exposure to the culture that was to become her nemesis.

During college, Gaillard also spent two summers at Brookhaven National Laboratory in Upton, New York, where she became hooked on high-energy particle physics. There, she met her first husband: Jean-Marc Gaillard, a postdoc at Columbia University in New York City. She did a graduate year at Columbia, then Jean-Marc was offered a post in Orsay near Paris. His colleagues advised Gaillard to accompany him, and to become "self-taught, like all great European

physicists". The first year in Orsay became "the worst year", as Gaillard "learnt to be a housewife" and was largely left on her own.

Jean-Marc was then offered a six-year staff position at CERN, the European centre for particle physics near Geneva, Switzerland. Here, Gaillard became a long-term visitor for some 20 years. Through Jean-Marc's connections, she secured space in a shared basement office in the CERN theory group; commuted between Orsay and CERN; and worked on the difference between matter and antimatter.

She was subjected to the "determined antifeminism" of theory-group leader Leon Van Hove, who became CERN director-general. He would ignore her and ask male colleagues about a project, for instance. Juggling research and her children, Gaillard became a major player in theoretical particle physics. Without the flexibility to interact with colleagues, she wrote many early papers alone. It is easy to sympathize when she tells of forgetting to collect her eight-year-old son Bruno from a music lesson in midwinter, or giving him the wrong bus fare so he was fined.

In 1973, Gaillard spent a "pivotal year" as a visitor at Fermilab near Chicago, Illinois, which was buzzing with excitement about the proposed theory of weak interactions. She met Benjamin Lee,

with whom she predicted the mass of the charm quark, gaining a "sort of star status". Back at CERN and working with giants such as John Ellis and Dimitri Nanopoulos, she turned to the decay modes of the Higgs boson, the signature for gluons and the mass of the bottom quark. She co-authored the paper that introduced the term penguin diagram for a type of loop-containing Feynman diagram (a description of interactions between subatomic particles). The same paper earned Bruno, then nine, an acknowledgement for "help with the calculations".

When Gaillard headed back to the United States in 1981 for the tenured position at Berkeley, it was not with Jean-Marc but with her soon-to-be second husband Bruno Zumino, a supersymmetry theorist who died last year. Gaillard became a grande dame of particle physics, with positions on many committees that shaped particle-physics research in the United States and, ultimately, the world.

The story is as much about a thrilling period in particle physics as about Gaillard's struggle to establish herself in a male-dominated sphere. Hers is the era of the standard model and its description of fundamental particles and forces. It has also seen the discovery of the Higgs boson, "a bug sloshing through molasses" as Gaillard describes it.

Gaillard explains her contributions clearly and without equations; exquisite illustrations by her son Bruno are reproduced. A fine example is the paper, presented as a conversation between herself, Lee and Jonathan Rosner, that used indirect experimental observations to predict the mass of the charm quark — three months before it was discovered (M. K. Gaillard *et al. Rev. Modern Phys.* 47, 277–310; 1975). Now 76, Gaillard continues to add to her broad portfolio, with a focus on superstrings and a desire to link theoretical predictions to experimental observation.

In 1980, Gaillard produced the *Report on Women in Scientific Careers at CERN*. This addressed the fact that just 3% of CERN staff were women, and called for the elimination of gender discrimination through equality in promotion, maternity leave and provision of a full-day crèche. Only in 1994 was a female experimental physicist, Fabiola Gianotti — who will become CERN director-general in 2016 — appointed to a senior position. Unfortunately, the current list of physicists in the CERN theory group shows no women in permanent senior positions. This is no reflection on Gaillard. As a colleague comments in the book: "She did it all!" ■

Val Gibson is an experimental particle physicist who has worked at CERN near Geneva, Switzerland. She is head of high-energy physics in the Cavendish Laboratory at the University of Cambridge, UK, and is a champion of equality and diversity. e-mail: gibson@hep.phy.cam.ac.uk

A Singularly Unfeminine Profession: One Woman's Journey in Physics

MARY K. GAILLARD
World Scientific: 2015.

Correspondence

Partner crop plants with solar facilities

About one million hectares of land will be required in the United States by 2030 to meet solar-energy targets (go.nature.com/2g5hkg). Cultivating carefully selected plants on such sites could offer a sustainable solution to meeting growing food and energy demands, particularly in regions with limited agricultural land and water resources (see, for example, go.nature.com/acixb7 and go.nature.com/n2sysg).

Photovoltaics (for producing electricity) and photosynthesis (for producing food, fodder or biofuel) both need sunlight. Large solar infrastructures protect vegetation from intense sun and strong winds, and regular washing of their surfaces provides water for the plants. Crops could be grown in the spaces between these structures to benefit from concentrated rainfall. These crops would reduce dust from disturbed soils, which could otherwise lower the efficiency of solar installations, and they would create extra revenue and employment.

The benefits and trade-offs of such co-located systems are now being evaluated (see, for example, go.nature.com/acixb7 and S. Ravi *et al. Environ. Sci. Technol.* **48**, 3021–3030; 2014). And solar operators and investors in North Africa, India, Mexico and the United States are already expressing an interest (S. R., personal communication). **Sujith Ravi** Temple University, Philadelphia, Pennsylvania, USA. sravi@temple.edu

STEM teaching: use more innovations

Two other concerns should be added to your prescriptions for improving teaching in science, technology, engineering and mathematics (STEM; see *Nature* **523**, 272–274 and 282–284; 2015).

We know that smaller class sizes and classrooms designed

for active learning give better academic outcomes (S. Cotner *et al. J. Coll. Sci. Teach.* **42**, 82–88; 2013), yet budgetary pressures discourage institutions from abandoning big lecture halls in favour of small classes.

Also, there should not be separate faculty tracks for teaching and research. Teaching positions rarely include research support, so they do not offer the same academic opportunities as research faculty positions.

These issues are ultimately about institutional and administrative buy-in. The success of STEM students depends on institutions investing in improved learning facilities and on administrators providing research, tenure and promotion opportunities for those who teach. **Luke Holbrook** Rowan University, Glassboro, New Jersey, USA. holbrook@rowan.edu

STEM teaching: avoid Swiss-cheese effect

You propose a shift from traditional university lectures to a system that teaches the methods of scientific enquiry to students of science, technology, engineering and mathematics (STEM; see *Nature* **523**, 272–274 and 282–284; 2015). This move has clear merits, but systematic transfer of the requisite knowledge should not be abandoned entirely.

A pioneer in active-learning practices, Roskilde University in Denmark has been using problem-based teaching and successfully involving students in research since 1972. Our experience shows, however, that there are potential pitfalls.

Unless critical thinking is allied with a strong fundamental knowledge base, there is a risk that students will develop a ‘Swiss cheese’ understanding of science — with a good grasp of their chosen subject areas but major gaps in others. This can produce niche researchers who lack a proper understanding of their wider field.

We therefore advise retaining aspects of traditional education in an appropriate balance, which is then adjusted on the basis of student and course evaluations.

Farhan R. Khan, Gary T. Banta Roskilde University, Denmark.
Christina Sørensen University of Oslo, Norway.
frkhan@ruc.dk

The future of public trust in science

The challenges of maintaining trust in science (see *Nature* **522**, 6; 2015) can be understood in terms of corrupting pressures that make it harder for scientists to do the good work to which many aspire.

The sheer scale of science today is destroying colleague communities; it also demands ‘objective’ metrics of quality, which are perverse and corruptible. These effects are compounded by imported commercial pressures. The idealism that motivated ‘little science’ is no longer plausible.

Maintaining the public’s trust in science calls for an urgent evaluation of its imperfections and vulnerabilities. We must identify what needs to be unlearned in the prevalent understanding of science: for example, we now know that any science-related policy problem poses more questions and solutions than can be derived from the illusory precision of models and indicators (a factor in the 2008 financial crisis).

Social-media channels are starting to teach the public more about new views of science. The growth of ‘DIY science’, which owes only minimal deference to established institutions, will eventually influence science education, and to good effect. In much the same spirit as citizen science has developed in parallel with established science, a movement of scientifically aware citizens could emerge within science. These citizens would develop an understanding of

the connection between science’s internal problems, such as morale and quality assurance, and external pressures of the sort we describe.

Jerome Ravetz University of Oxford, UK.

Andrea Saltelli University of Bergen, Norway.
jerome.ravetz@gmail.com

Solar ovens beaten by rain and tortillas

Solar ovens sometimes fall short of their promise as the gold standard of clean cooking, despite producing zero emissions (see L. S. Brown and W. F. Lankford *Nature* **521**, 284–285; 2015).

In a solar-oven project funded by the Central American Solar Energy Project (CASEP) in Nicaragua, participants generally reported large fuel savings. Yet objective measurements found that savings were not significant, and surveys indicated that users continued to cook on biomass-burning stoves. Data from thermometers inside solar ovens confirmed that oven usage was widely over-reported (see go.nature.com/oqff6a).

Furthermore, although the 77 Nicaraguan women interviewed in 2014 (by S.V.) found CASEP’s empowerment training helpful, this did not result in self-sustaining community action after the project ended.

In parts of Latin America, preparing tortillas accounts for more than half of cooking fuel usage (O. Masera *et al. Energy Sustain. Dev.* **11**, 45–56; 2007), but tortilla preparation is impossible with most solar-cooker designs because they are not hot enough. And in areas with a rainy season, solar cooking is impractical for half of the year. Improved biomass stoves and biogas might be more effective solutions in such regions.

Gordon Bauer University of Oslo, Norway.

Sarah Vukelich Williams College, Williamstown, Massachusetts, USA.
gordon.bauer@gmail.com

ORGANIC CHEMISTRY

A cure for catalyst poisoning

Compounds that are sensitive to the components of air are difficult to use in chemical reactions, requiring conditions that are tedious to set up. A simple, practical solution to this problem has finally been devised. [SEE LETTER P.208](#)

MARCUS E. FARMER & PHIL S. BARAN

Capsules or pills for drug delivery were invented by the French pharmacists François Mothes and Joseph Dublanc in the early 1800s^{1,2} as a reproducible method for the consistent dosing and delivery of medicines and vitamins, and to maintain the stability of pharmaceutically active ingredients. This remains the most common formulation mode for drugs — without it, pharmacists would have to carefully weigh out and dispense freshly prepared powders of drug substances to patients. Yet research chemists still have to do this for each compound used in their reactions. This is especially problematic when using reagents and catalysts that are sensitive to atmospheric water vapour, oxygen or carbon dioxide. On page 208 of this issue, Buchwald and colleagues³ describe an ingenious solution to this problem by ‘formulating’ sensitive compounds in capsules, thereby eliminating the inconvenience associated with their storage and handling.

Large and expensive sterilized boxes — known as dry boxes or glove boxes — containing chemically inert gases are widely used to prevent sensitive catalysts and reagents from being exposed to components of the atmosphere (Fig. 1). Dry boxes have enabled many useful discoveries for chemical synthesis in academic settings, but the industrial application of discoveries made in such conditions is hampered by the inconvenience associated with assembly, maintenance and labour-intensive operation. Many useful and enabling chemical transformations thus remain under-used, or are completely ignored. It is surprising that this glaring unmet chemical need has not previously been addressed.

The largest group of consumers for new chemical methods work in the pharmaceutical, agrochemical and materials sectors. Buchwald and co-workers therefore focused on developing a technique that would bring glove-box chemistry to the open bench, where the chemists from these industries feel most comfortable. To ensure the successful adoption of their technique, they set out to develop an approach that could easily transition a variety of synthetic methods from the glove box to the bench top without the need to substantially

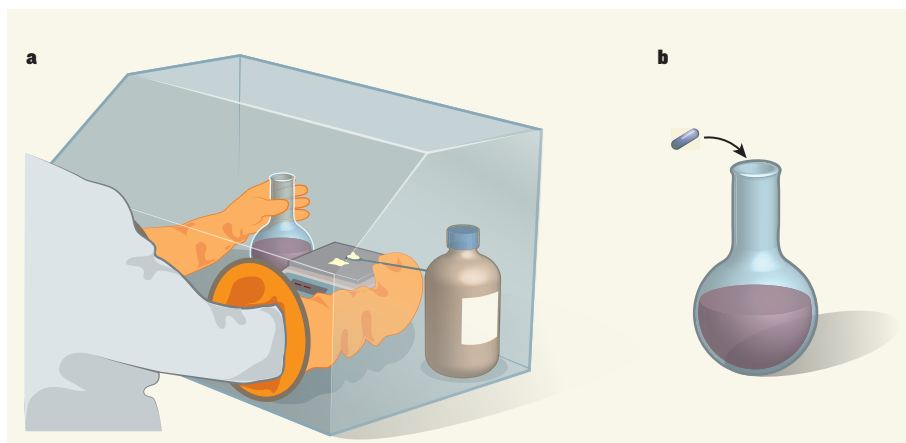


Figure 1 | Encapsulated reagents for dispensing air-sensitive compounds. **a**, Research chemists typically store, manipulate and perform reactions involving air-sensitive reagents in glove boxes — sterilized chambers that contain a chemically inert atmosphere. This is practically much more difficult than performing reactions on the open bench. **b**, Buchwald and colleagues³ report that air-sensitive reagents can be conveniently used and stored on the open bench when sealed in capsules made from paraffin wax, which melt on heating.

alter pre-existing reaction conditions.

The authors were initially inspired by the example of potassium hydride — a moisture-sensitive reagent that is sold as a dispersion in paraffin wax, allowing it to be stored and manipulated without using a glove box. Buchwald and colleagues therefore tried to form dispersions of reactants using molten paraffin wax, but obtained random, irreproducible distributions of the compounds within the resulting mixtures.

To circumvent this problem, the researchers developed a technique that places reagents and catalysts in a paraffin-wax capsule. The capsules can be stored in the open air for months, and retain full activity even when dipped in water. Remarkably, the researchers observed that 2-pyridylzinc chloride dioxanate — an air-sensitive reagent that degrades in minutes to hours if not protected — can be stored on the bench top for more than a year without signs of degradation if it is sealed in a paraffin capsule. Furthermore, because paraffin wax is generally unreactive, the capsules can simply be added directly to reaction mixtures using common laboratory procedures. The capsule melts on heating, releasing its contents, and the molten wax does not interfere with the desired chemical reaction.

To showcase the utility of this approach, Buchwald and co-workers used their capsules in synthetically useful reactions known as nucleophilic fluorinations of aryl triflates^{4,5} (see Fig. 1b of the paper³). These reactions require an oxygen-sensitive palladium catalyst and caesium fluoride (which is highly hygroscopic — that is, it rapidly absorbs moisture from the atmosphere). Exposure of these compounds to air usually sidetracks the reaction: oxygen causes the catalyst to be poisoned (to lose its activity), whereas moisture causes the formation of unwanted byproducts. The authors therefore encapsulated a mixture of these compounds, added the capsules to reactions set up using standard laboratory conditions, and compared the outcomes with the same reactions run in a glove box^{4,5}. In all cases, the reactions provided comparable results, indicating that the encapsulated reagents had been successfully protected from air.

To further demonstrate the feasibility of their approach to enable the bench-top assembly of palladium-catalysed fluorination reactions, the researchers prepared capsules containing a mixture of three air-sensitive compounds that fluorinates a variety of substrates called aryl and heteroaryl bromides⁶ (see Fig. 2 of the paper³). As with the previous fluorination

procedure, the yields obtained on the bench top using the prepared capsules provided similar yields to those obtained in the glove box.

Next, Buchwald and colleagues showcased their technique for a carbon–nitrogen bond-forming reaction (an amination; see Fig. 3 of the paper³) that they had previously developed in their laboratory and that is now widely used in every branch of chemistry⁷. The reaction requires a palladium precatalyst (a palladium compound that is converted to an active catalyst during a reaction) and a strong hygroscopic base. The authors encapsulated the precatalyst and the base together, and found that the resulting mixture was stable when stored on a bench top for more than eight months. The result demonstrates that the compounds can coexist under these conditions even though the precatalyst is activated by bases when in solution. This ‘amination capsule’ worked as well on the bench top as reactions tediously prepared in a glove box⁸.

As a finale, Buchwald and co-workers showed that the encapsulation technique could enable carbon–carbon bond-forming reactions known as Negishi cross-couplings, which use moisture-sensitive zinc reagents (see Fig. 4 of the paper³). The encapsulation of one such reagent, 2-pyridylzinc chloride dioxanate, with an appropriate precatalyst facilitated such reactions on the open bench in comparable yields to the analogous glove-box procedure⁹.

Although the encapsulation approach carries many benefits, it will not completely eliminate the glove box because the capsules still need to be prepared in the absence of air. The widespread adoption of this approach for synthesis will also rely on the availability of the capsules, although we expect that vendors will expedite their commercialization and distribution.

The thought-provoking technique opens up many avenues for exploring the reactivity of air-sensitive reagents and catalysts, both in academic settings and in the many areas of industrial science that require the rapid and automated preparation of libraries of structurally diverse compounds. If many catalysts and reagents become readily available as capsules, the influence of this approach will probably be seen in the pharmaceutical, agricultural and materials industries. It may not be too unrealistic to predict that these capsules will do for organic chemistry what Mothes and Dublanc’s pills did for medicine. ■

Marcus E. Farmer and Phil S. Baran are in the Department of Chemistry, Scripps Research Institute, La Jolla, California 92037, USA.
e-mail: pbaran@scripps.edu

1. Mothes, F. A. B. French patent 9690 (1834).
2. Wilbert, M. I. *Am. J. Pharm.* **85**, 559–572 (1913).
3. Sather, A. C., Lee, H. G., Colombe, J. R., Zhang, A. & Buchwald, S. L. *Nature* **524**, 208–211 (2015).
4. Watson, D. A. *et al. Science* **325**, 1661–1664 (2009).

5. Lee, H. G., Milner, P. J. & Buchwald, S. L. *Org. Lett.* **15**, 5602–5605 (2013).
6. Lee, H. G., Milner, P. J. & Buchwald, S. L. *J. Am. Chem. Soc.* **136**, 3792–3795 (2014).
7. Ishihara, Y., Montero, A. & Baran, P. S. *The Portable Chemist’s Consultant: A Survival Guide for Discovery,*

- Process, and Radiolabeling* (Apple, 2013).
8. Fors, B. P. & Buchwald, S. L. *J. Am. Chem. Soc.* **132**, 15914–15917 (2010).
9. Colombe, J. R., Bernhardt, S., Stathakis, C., Buchwald, S. L. & Knochel, P. *Org. Lett.* **15**, 5754–5757 (2013).

REGENERATIVE BIOLOGY

Maintaining liver mass

A previously under-appreciated subset of liver cells has been found to contribute to the day-to-day maintenance of liver mass in mice. The cells are induced and supported by signals from an adjacent vein. SEE ARTICLE P180

KENNETH S. ZARET

The liver has remarkable regenerative powers. Many studies have focused on the ability of different types of cell to replenish both the liver and the bile ducts after damage¹, but less clear is how the liver self-renews when cells die naturally. Such homeostatic renewal ensures that the liver maintains an appropriate mass, and so is crucial for health. On page 180 of this issue, Wang *et al.*² shed light on this issue, focusing on an under-appreciated, self-renewing cell population in the undamaged livers of mice. It seems that liver cells themselves might function as ‘stem cells’ for homeostasis when they are positioned in a specialized zone of the liver.

After food has been eaten, nutrients, along with any toxins that have been ingested, are absorbed by the intestine, pass into the bloodstream and are transported directly to the liver

for metabolic processing. Liver cells (hepatocytes) control metabolism and act as the first line of defence against toxins. But hepatocytes can be damaged in the line of duty, and chronic liver damage is a major health concern worldwide. Identifying the cell populations in the liver that can repair damage has therefore been a topic of intensive research.

Consider the vascular plumbing of the liver. Nutrient-rich blood from the intestines travels along the portal vein, arriving in the portal zone of the liver, where bile ducts and the hepatic artery also reside. The blood then courses through sinuses in the liver mass, is exposed to hepatocytes for metabolite and toxin exchange, and collects in the central vein. Thus, the portal zone must contend with greater toxic insults than the central zone. Indeed, periportal hepatocytes respond to most forms of liver damage^{1,3–5} and can also contribute to homeostatic cell renewal^{1,3}.

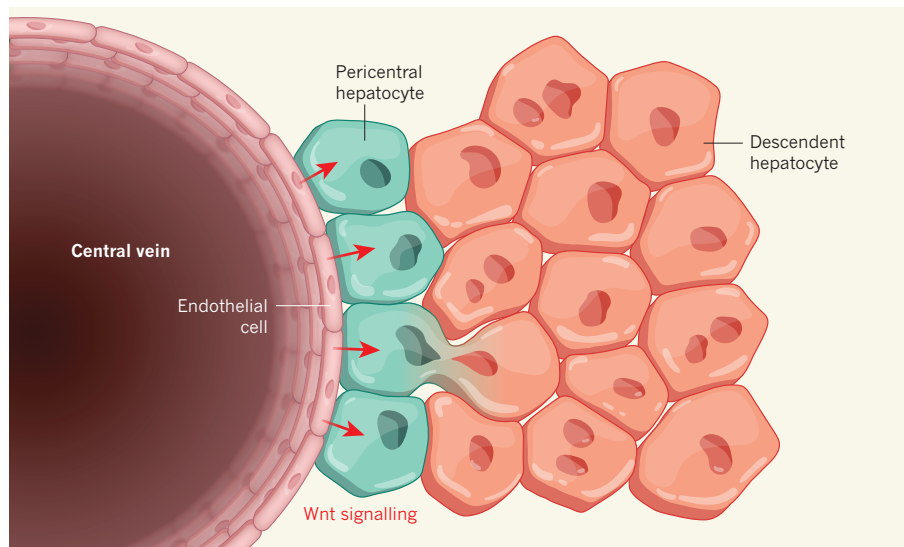


Figure 1 | A contributor to everyday liver regeneration. The endothelial cells lining the central vein of the liver emit Wnt signals that induce the expression of Wnt-responsive genes in adjacent pericentral liver cells (hepatocytes). Wang *et al.*² report that these signals also stimulate the proliferation of pericentral hepatocytes. The cells give rise to descendants that reside beyond the reach of Wnt signals, and that replicate more slowly than their parents (some of the descendants have more than one nucleus). In this way, pericentral hepatocytes contribute to the maintenance of liver mass.

However, the liver's anatomy indicates that the central zone might constitute a more-protected reservoir of cells, and might be a preferable location for cells involved in homeostatic self-renewal.

Hepatocytes in the pericentral region immediately adjacent to the central vein are known to replicate slightly faster than other hepatocytes in normal conditions⁶, and to be the only hepatocyte population that expresses genes activated by the Wnt signalling pathway^{7,8}. Wang and colleagues used genetic techniques in mice to indelibly label cells expressing a Wnt-responsive gene, such that these cells and their descendants fluoresced. They then tracked this fluorescent lineage and showed that pericentral hepatocytes self-renew — the cells remain close to the central vein and are not normally replaced by other hepatocytes. Over time, these cells give rise to descendants outside the pericentral zone that can replenish up to 40% of the liver's mass under normal conditions (Fig. 1). These findings lead to the question of how pericentral hepatocytes differ from other hepatocytes, and whether such differences depend on proximity to the central vein. Consistent with this, it is now recognized that a stem cell's identity can be dependent on the signals that it receives from its local environment^{9,10}.

Mammalian cells typically carry two copies of each chromosome, but most hepatocytes carry several copies of this chromosome complement and exhibit chromosomal imbalances on division, making them less than ideal candidates for the population that replenishes the liver¹¹. Wang *et al.* observed that many pericentral hepatocytes have the normal chromosome complement, and so seem better suited to replicating their genomes faithfully when they divide. Finally, the authors found that Wnt signals released from the endothelial cells that make up the central vein are required to maintain the proliferation of pericentral hepatocytes and thus their function in replenishing liver cells.

The discovery that pericentral hepatocytes, along with other hepatocytes^{1,3}, contribute to liver homeostasis opens up many avenues for study. For instance, the relative contribution of each of these cell types to homeostatic regeneration is not known. The role of pericentral hepatocytes in regeneration following non-periportal forms of liver damage also remains to be determined. Could a better understanding of the cells enable self-renewal to be enhanced? Manipulating the Wnt pathway *in vivo* might provide insights along these lines, as has been shown for liver 'organoids' grown *in vitro*¹².

Perhaps the most important question is whether the pericentral hepatocytes behave as a niche-dependent stem-cell population^{9,10}; that is, whether any hepatocyte placed in the pericentral region, under the influence of endothelial Wnt signalling, would become

Wnt-responding, faster-replicating cells, functioning like the original pericentral cells. This crucial test could be carried out by ablating pericentral cells, for example through the transient induction of diphtheria toxin, and determining whether other hepatocytes take on their role. If it could be shown that any hepatocyte — or at least any hepatocyte with a normal chromosome complement — when placed in the pericentral region or exposed to the correct Wnt signalling could be 'activated' to become a more-efficient cell for liver homeostasis, this could have an impact on treatments for chronic liver disease.

But almost all hepatocytes, regardless of their position in the liver, can self-renew and contribute to liver homeostasis^{1,3}. Thus, it may be that it is not appropriate to consider whether any one hepatocyte population is the true homeostatic stem cell. Instead, a more pertinent question might be whether some hepatocytes are better at self-renewing than others.

Interestingly, the authors found that pericentral hepatocytes are the only adult hepatocyte population to express Tbx3, a transcription factor that is essential for the development of hepatoblasts¹³, the precursors of hepatocytes and bile-duct cells in the early embryo. Direct signalling from adjacent endothelial cells promotes embryonic hepatoblast growth¹⁴. Thus, the pericentral hepatocytes live in an environment that shares features with embryonic liver development. But hepatoblasts are bipotential, whereas the pericentral hepatocytes seem to

give rise to hepatocytes only, indicating differences in the networks that regulate these cell types. Understanding the similarities and differences between the pericentral hepatocytes and hepatoblasts, and between pericentral hepatocytes and other hepatocytes in the liver, is sure to provide crucial insights for the liver and regeneration research fields. ■

Kenneth S. Zaret is at the Institute for Regenerative Medicine and Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. e-mail: zaret@mail.med.upenn.edu

1. Miyajima, A., Tanaka, M. & Itoh, T. *Cell Stem Cell* **14**, 561–574 (2014).
2. Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. *Nature* **524**, 180–185 (2015).
3. Malato, Y. J. *Clin. Invest.* **121**, 4850–4860 (2011).
4. Tarlow, B. D. *et al. Cell Stem Cell* **15**, 605–618 (2014).
5. Yanger, K. *et al. Cell Stem Cell* **15**, 340–349 (2014).
6. Magami, Y. *et al. Liver* **22**, 419–425 (2002).
7. Benhamouche, S. *et al. Dev. Cell* **10**, 759–770 (2006).
8. Yang, J. *et al. Hepatology* **60**, 964–976 (2014).
9. Rompolas, P., Mesa, K. R. & Greco, V. *Nature* **502**, 513–518 (2013).
10. Sheng, X. R., Brawley, C. M. & Matunis, E. L. *Cell Stem Cell* **5**, 191–203 (2009).
11. Duncan, A. W. *et al. Nature* **467**, 707–710 (2010).
12. Huch, M. *et al. Nature* **494**, 247–250 (2013).
13. Suzuki, A., Sekiya, S., Büscher, D., Izpisua Belmonte, J. C. & Taniguchi, H. *Development* **135**, 1589–1595 (2008).
14. Matsumoto, K., Yoshitomi, H., Rossant, J. & Zaret, K. S. *Science* **294**, 559–563 (2001).

This article was published online on 5 August 2015.

DNA REPLICATION

Strand separation unravelled

The DNA double helix must be separated into single strands to be duplicated. A structure of the Mcm2–7 helicase enzyme responsible for this activity yields unprecedented insight into how the process is initiated. SEE ARTICLE P. 186

**MATTHEW L. BOCHMAN
& ANTHONY SCHWACHA**

The successful replication of double-stranded DNA, an essential part of cell division, depends on a helicase enzyme that separates the two component strands. Although simple helicases have been extensively studied¹, much less is known about the complex replicative helicases found in eukaryotes (the group of organisms that includes animals, plants and fungi). But that is about to change. On page 186 of this issue, Li *et al.*² capitalize on advances in cryo-electron microscopy³ to resolve the structure of a eukaryotic

helicase, Mcm2–7, to a near-atomic resolution of 3.8 ångströms — around five times higher than the best Mcm2–7 structure reported so far⁴. Combined with previous studies, this structure indicates how a key step in DNA replication occurs: the initial 'melting' of double-stranded DNA into single strands.

Mcm2–7 has a central role in eukaryotic DNA replication. Like similar helicases from bacteria, archaea and viruses, it unwinds double-stranded DNA (dsDNA) by binding one strand in its central channel, excluding the other. Energy, provided by the enzyme's ability to hydrolyse ATP molecules, enables the complex to translocate along the bound DNA,

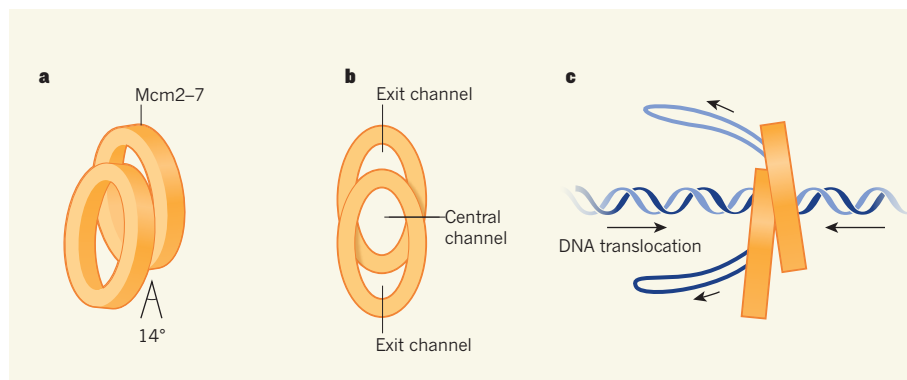


Figure 1 | Structure of a strand separator. **a**, The Mcm2–7 helicase is a doughnut-shaped enzyme composed of six different subunits (individual subunits not shown) that is vital for DNA replication⁵. Li *et al.*² resolve a structure that contains two copies of this hexamer, which they suggest is involved in the DNA melting process that converts double-stranded DNA (dsDNA) into single strands to initiate DNA replication. In the authors' structure, the two hexamers are tilted at a 14° angle relative to one another. **b**, This conformation forms a narrow central channel between the two hexamers through which dsDNA can pass, together with two exit channels. **c**, dsDNA pumped through the central channel might be extruded through the exit channels as single-stranded DNA 'rabbit ears', which can then be replicated.

resulting in unwinding of the complementary strand¹. But the doughnut-shaped Mcm2–7 is structurally and functionally different from other helicases, because it is the only known hexameric helicase to be derived from six different subunits (Mcm2 to Mcm7) instead of from six copies of the same subunit. This feature has allowed portions of the complex to evolve extra, specialized functions that are thought⁵ to be crucial to the enzyme's ability to load onto DNA and to activate its unwinding activity — two landmark regulatory events during DNA replication.

Two structures containing Mcm2–7 have been described previously. One represents the CMG complex^{6,7}, which is active during the phase of DNA replication known as elongation, when complementary DNA is synthesized for each existing strand. The complex contains one Mcm2–7 hexamer and two other essential replication factors that activate the enzyme's DNA unwinding ability. By contrast, the second structure^{8,9} is an inactive form of the enzyme, which has been isolated from cells before they replicate. This structure contains two Mcm2–7 hexamers in a head-to-head orientation, enclosing dsDNA in the central channel. Helicase structures such as this Mcm2–7 double hexamer (Mcm2–7 DH) are rare, and so its purpose has been a cause for debate.

One reasonable conjecture is that the Mcm2–7 DH participates in DNA melting. Whereas DNA unwinding enlarges a pre-existing single-stranded DNA (ssDNA) region during elongation, DNA melting, which is an earlier process, initiates replication by locally transforming dsDNA into ssDNA. Local melting provides a site for the subsequent assembly of a DNA replication fork — the full complement of proteins that enable duplication of the genetic material¹⁰. Although melting has been well studied in bacteria, little is known about how it occurs in eukaryotes¹⁰. Li and colleagues'

structure, when combined with other data, is highly consistent with a role for the Mcm2–7 DH in DNA melting, for several reasons.

First, both the current study and a previous one⁴ demonstrate that the two Mcm2–7 hexamers in the Mcm2–7 DH are offset along the long vertical axis of the hexamer, at a 14° tilt relative to one another (Fig. 1a). This offset restricts the dimensions of the central channel (Fig. 1b). Although DNA is not visible in the authors' structure, these data suggest that dsDNA will be kinked at the interface between the two hexamers. Sharp DNA bending is known to cause local DNA melting¹¹, and may contribute to the unwinding of dsDNA during transcription¹². Thus, a DNA kink between the two Mcm2–7 hexamers could serve to initiate DNA melting.

Second, although helicases normally interact productively only with ssDNA, a specific form of the Mcm complex (Mcm467) has been shown to bind to and translocate along dsDNA¹³. This is consistent with a potential role for Mcm2–7 in manipulating dsDNA during melting. Finally, unlike bacterial hexameric helicases, some viral replicative helicases, such as papillomavirus E1 and simian virus-40 (SV40) large T-antigen, initially form dsDNA-containing DHs that resemble the Mcm2–7 DH (refs 14, 15). These structures locally melt DNA and then uncouple into single hexamers to unwind DNA during elongation.

How might dsDNA melting occur? Electron microscopy indicates that the SV40 large T-antigen DH can act as a pump¹³, in which dsDNA enters each hexamer from flanking regions and ssDNA is extruded in 'rabbit ear' structures at the interface between them¹⁴. Consistent with such a mechanism in eukaryotes, the misalignment of the two hexamers in the Mcm2–7 DH creates two exit channels at the hexamer interface through which rabbit ears might be extruded (Fig. 1c).

Thus, the Mcm2–7 DH might melt DNA in a manner analogous to melting on SV40 large T-antigen, with local unwinding of the bent DNA forming a highly flexible hinge to facilitate ssDNA extrusion. Such a model had been proposed to explain Mcm2–7 DNA unwinding during elongation¹³. Because the Mcm2–7 DH seems to be enzymatically inactive, further research will be needed to identify the factors required to activate the DH for melting, as well as to determine how the individual Mcm2–7 hexamers physically uncouple and are remodelled into the ssDNA-bound form needed for elongation.

Given the technical advances in cryo-electron microscopy, a flood of high-resolution structures should become available in the near future. However, such structures provide only a static glimpse of the target protein, a particularly limiting problem for the study of dynamic processes such as DNA replication. Because Mcm2–7 is only one of many molecular motors involved in DNA replication, understanding the dynamic nature of their interactions is essential for a complete understanding of DNA replication. To this end, single-molecule studies using reconstituted eukaryotic replication systems^{7,16} have begun to shed much-needed light on the dynamics of this process. Together, these varied experimental approaches should yield a holistic understanding of the vital process of DNA replication. ■

Matthew L. Bochman is in the Molecular and Cellular Biochemistry Department, Indiana University, Bloomington, Indiana 47405, USA.

Anthony Schwacha is in the Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.
e-mails: bochman@indiana.edu; schwacha@pitt.edu

1. Enemark, E. J. & Joshua-Tor, L. *Curr. Opin. Struct. Biol.* **18**, 243–257 (2008).
2. Li, N. *et al.* *Nature* **524**, 186–191 (2015).
3. Kühlbrandt, W. *eLife* **3**, e03678 (2014).
4. Sun, J. *et al.* *Genes Dev.* **28**, 2291–2303 (2014).
5. Bochman, M. L. & Schwacha, A. *Microbiol. Mol. Biol. Rev.* **73**, 652–683 (2009).
6. Ilves, I., Petojevic, T., Pesavento, J. J. & Botchan, M. R. *Mol. Cell* **37**, 247–258 (2010).
7. Fu, Y. V. *et al.* *Cell* **146**, 931–941 (2011).
8. Evrin, C. *et al.* *Proc. Natl Acad. Sci. USA* **106**, 20240–20245 (2009).
9. Remus, D. *et al.* *Cell* **139**, 719–730 (2009).
10. Gai, D., Chang, Y. P. & Chen, X. S. *Curr. Opin. Struct. Biol.* **20**, 756–762 (2010).
11. Shin, J., Lee, O.-C. & Sung, W. J. *Chem. Phys.* **142**, 155101 (2015).
12. Tang, G.-Q. & Patel, S. S. *Biochemistry* **45**, 4936–4946 (2006).
13. Kaplan, D. L. & O'Donnell, M. *Mol. Cell* **15**, 453–465 (2004).
14. Schuck, S. & Stenlund, A. *Mol. Cell* **43**, 776–787 (2011).
15. Wessel, R., Schweizer, J. & Stahl, H. J. *Virology* **66**, 804–815 (1992).
16. Tica, S., Friedman, L. J., Ivica, N. A., Gelles, J. & Bell, S. P. *Cell* **161**, 513–525 (2015).

This article was published online on 29 July 2015.

Matter and antimatter scrutinized

A search for differences in the charge-to-mass ratio of protons and antiprotons, conducted at unprecedented levels of precision, results in stringent limits to the validity of fundamental physical symmetries. [SEE LETTER P.196](#)

KLAUS P. JUNGMAUN

The standard model¹ of particle physics is considered to be the best physical theory that we have. It is built on symmetries and can describe all the experiments and observations concerning the known subatomic particles. However, the model includes some 30 free parameters and is not fully explanatory. For example, it cannot explain a profound mystery of physics and cosmology², the fact that there is no antimatter in the Universe. When matter and antimatter mutually annihilated each other following the Big Bang, any pre-existing symmetry between them was broken. Matter but no antimatter was left behind, and we lack a satisfactory explanation as to how this occurred³. Research on the fundamental differences between particles and antiparticles may provide an answer. In this vein, Ulmer *et al.*⁴ (page 196) perform a high-precision, comparative study of the properties of protons and antiprotons.

The authors used negatively charged

hydrogen atoms (which represent protons for technical reasons) and individual antiprotons, the latter generated by the antiproton decelerator facility at CERN, Europe's particle-physics laboratory near Geneva, Switzerland. These species were stored in a sophisticated device known as a Penning trap, which consists of metal electrodes placed at defined electric potentials inside a strong and stable magnetic field (Fig. 1). In the trap, which has a diameter of just a few millimetres, the motion of electrically charged particles is similar to that in an accelerator such as the Large Hadron Collider at CERN, but the energies attained are 10^{15} times smaller.

A particle's cyclical motion in the Penning trap has a characteristic frequency (known as the cyclotron frequency), which is proportional to the magnetic field strength and the particle's charge-to-mass ratio. Ulmer *et al.* determined the cyclotron-frequency ratio for the antiproton and the negative hydrogen ion, alternately recycling the same individual particles at intervals of a few minutes from each

other in the same experiment. The authors repeated this procedure 6,500 times within 35 days and scrutinized the results for systematic errors. Finally, they found that the charge-to-mass ratios of protons and antiprotons are equal to within 69 parts per trillion.

This result is four times more accurate than previous measurements⁵ of these ratios, and has implications for the validity of fundamental physical symmetries and theories that have been proposed to address unexplained aspects of the standard model. Symmetries have a central role in physics. A symmetry that holds across the Universe is an indication that a conservation law is at work. For example, adjusting a clock by an arbitrary time interval leaves all physical processes completely unaffected. A consequence of this is that energy can neither be created nor destroyed. But, when a symmetry is violated or a quantity is not conserved, a symmetry-breaking process must be at work.

In the process known as nuclear β -decay, for instance, a neutron is transformed into a proton, an electron and an antineutrino, but only antineutrinos of 'right-handed' nature appear. As a consequence, the electron is emitted into a preferred direction with respect to the neutron spin. This asymmetry is an example of parity (P) violation⁶, which means that β -decay would not proceed in exactly the same way in a mirrored version of the world. Similar symmetry violations are observed only in some processes that involve the weak force. They can appear if the signs of electric charges are reversed (charge conjugation, C), or if the arrow of time changes direction (time reversal, T). Symmetry violations also occur when the combination of C and P symmetries (CP symmetry) breaks down; these become evident for physical processes that occur differently when the signs of charges and handedness are changed simultaneously.

The physicist Andrei Sakharov offered⁷ an explanation for the observed dominance of matter, based on such a CP-symmetry violation. However, all the known CP-violating processes cannot sufficiently explain the preponderance of matter over antimatter. Furthermore, at current levels of precision, no physical process has been found to violate the combination of C, P and T symmetries (CPT symmetry), which relates to fundamental physical principles. In quantum mechanics, for example, this combined symmetry ensures that particle spins take only integer and half-integer values. Moreover, the invariance of physical laws in different moving frames of reference (known as the Lorentz invariance) implies CPT symmetry^{8,9}.

Physicist Alan Kostelecký and colleagues have suggested that a violation of this symmetry might provide an alternative explanation for the missing antimatter¹⁰. Unlike Sakharov's model, which requires the disappearance of antimatter in the early, thermally unstable Universe, the latter model does not

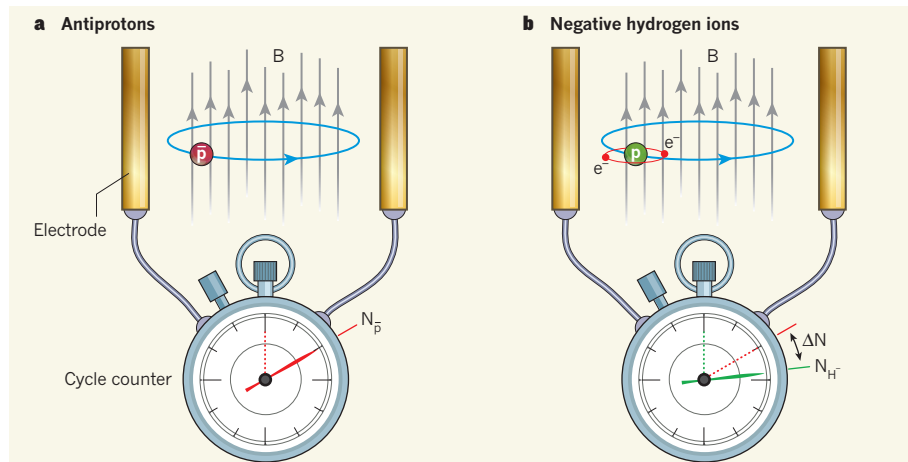


Figure 1 | Particle and antiparticle motion. Ulmer *et al.*⁴ used a device known as a Penning trap to measure, under identical conditions, the characteristic cycling frequency of **a**, antiprotons (\bar{p}) and **b**, negatively charged hydrogen ions (H^- , in lieu of protons; represented as a proton (p) and two electrons (e^-)) undergoing circular motion in a magnetic field of strength B (grey arrows), set perpendicular to the direction of motion. From the cycling frequency, which is the number of cycles ($N_{\bar{p}}$ and N_{H^-}) that each particle type completed per unit of time, the charge-to-mass ratios of pairs of individual antiprotons and negatively charged hydrogen ions were determined. The number of cycles was measured from signals registered by the trap's electrodes. After correcting for the difference (ΔN) between $N_{\bar{p}}$ and N_{H^-} to take into account the binding energies and the masses of the two electrons in H^- that render it different from a proton, the authors found that the charge-to-mass ratios of protons and antiprotons are identical with an accuracy of 69 parts per trillion.

have this additional stringent condition. Under CPT symmetry, particles and antiparticles are strictly identical except for the sign of their charge. Ulmer and colleagues' measurements of the proton and antiproton charge-to-mass ratios place limits on the differences between the properties of particles and antiparticles and establish a tighter boundary on a possible CPT-symmetry violation.

The charge-to-mass ratios measured by the authors do not vary by more than 720 parts per trillion during a sidereal day, which is the duration of a day with respect to the fixed positions of stars rather than to the Sun. Therefore, this level of accuracy excludes a violation of the CPT symmetry or of the related Lorentz invariance that could be attributed to a preferred frame of reference, such as the one provided by the cosmological microwave background (the Big Bang's relic radiation). It should also be noted that because the cyclotron frequency measurements took place in Earth's gravitational field, any difference in the way that protons and antiprotons interact with gravity would modify their respective frequencies¹¹. However, the authors found no such difference larger than 870 parts per billion. This means that the weak equivalence principle — which states that all bodies in a given gravitational field undergo the same acceleration independently of their properties — holds at this level of accuracy.

Ulmer and colleagues' experiment has improved our understanding of fundamental physical principles by placing important limits on several processes. This experiment is a highlight of research on the central question of the prevailing matter–antimatter asymmetry, which the researchers approach by a promising route. Apart from the authors' tests of the CPT-symmetry invariance, there are other experiments¹² that have searched for violations of the CP and T symmetries. The search for the former typically involved precise measurements of particle properties, including antiprotonic systems. The hunt for the latter included searches for the elusive permanent electric dipole moments of particles, and research on the correlations in the parameters of β -decaying nuclei and their decay products, such as neutrinos, electrons and daughter nuclei.

Highly precise experiments at low energies, such as this, are complementary to searches for evidence of fundamental symmetry violations in high-energy particle colliders. There is still no indication whether CPT- or CP-symmetry violations may be responsible for the matter–antimatter asymmetry and for any possible, but as yet unknown, differences between particles and antiparticles. Scientists therefore look forward to improved results from ongoing, well-motivated precision experiments³, involving antiprotons in particular¹³, which sustain the attack on one of the most intriguing questions in physics. ■

Klaus P. Jungmann is at the Van Swinderen Institute for Particle Physics and Gravity, University of Groningen, 9447AA Groningen, the Netherlands.

e-mail: k.h.k.j.jungmann@rug.nl

1. Olive, K. A. *et al.* *Chinese Phys. C* **38**, 090001 (2014).
2. Dolgov, A. D. *Frascati Phys. Ser.* **XXXVI**, 507–516 (2004).
3. Willmann, L. & Jungmann, K. *Ann. Phys. (Leipz.)* <http://dx.doi.org/10.1002/andp.201500008> (2015).
4. Ulmer, S. *et al.* *Nature* **524**, 196–199 (2015).

MOLECULAR BIOLOGY

It takes two to untangle

Yeast require the enzyme Hsp104 to untangle protein aggregates, which arise in stressed or aged cells. Animals lack Hsp104, but it emerges that proteins of the DNAJ family of molecular chaperones can fulfil this role. [SEE LETTER P.251](#)

HARM H. KAMPINGA

The intracellular protein quality-control network ensures that proteins fold properly or are soon degraded when damaged or no longer needed¹. When the quality control fails, proteins can clump together in aggregates — a phenomenon associated with stress and ageing, and with many neurodegenerative diseases, several cardiac- and skeletal-muscle diseases and diabetes type II (ref. 2). In yeast, a molecular chaperone called heat-shock protein 104 (Hsp104) mediates disaggregation³, thus maintaining cellular health. But although several observational studies have suggested that animals have the potential to disaggregate proteins⁴, they lack a functional Hsp104 equivalent. Furthermore, *in vitro* disaggregation using human molecular chaperones has proved inefficient⁵. On page 251 of this issue, Nillegoda *et al.*⁶ show that protein disaggregation in animals is mediated by synergistic cooperation between different members of another class of molecular chaperone, the DNAJ proteins.

In both yeast and animals, Hsps prevent the formation of aggregates by binding to hydrophobic stretches of amino acids. Small Hsps, together with DNAJ proteins, capture unfolded or misfolded proteins and maintain them in a soluble state. These captured clients can then be transferred to proteins of the Hsp70 family, which mediate refolding or degradation, thus preventing aggregation¹. DNAJ proteins — the largest group of molecular chaperones, with 22 members in yeast and more than 50 in humans — are thought to play their part in this process by directing Hsp70 to specific clients⁷. In addition, individual DNAJ family members assist Hsp70 in Hsp104-dependent

5. Gabrielse, G. *et al.* *Phys. Rev. Lett.* **82**, 3198–3201 (1999).
6. Wu, C. S., Ambler, E., Hayward, R. W., Hoppes, D. D. & Hudson, R. P. *Phys. Rev.* **105**, 1413–1415 (1957).
7. Sakharov, A. D. *JETP Lett.* **5**, 24–27 (1967).
8. Schwinger, J. *Phys. Rev.* **82**, 914–927 (1951).
9. Lüders, G. *Ann. Phys.* **2**, 1–15 (1957).
10. Bertolami, O., Colladay, D., Kostelecký, V. A. & Potting, R. *Phys. Lett. B* **395**, 178–183 (1997).
11. Hughes, R. J. & Holzschneider, M. H. *Phys. Rev. Lett.* **66**, 854–857 (1991).
12. Blaum, K., Müller, H. & Severijns, N. (eds) *Ann. Phys. (Leipz.)* **525**, A127–A143, 539–737 (2013).
13. Johansson, T., Froehlich, P. & Jonsell, S. (eds) *Proc. LEAP 2013 Hyperfine Interact.* **228**, 1–165 (2014).

protein disaggregation in yeast. However, the role of DNAJ proteins in the solubilization of aggregates in animals has been enigmatic.

DNAJ proteins are divided into A, B and C classes, of which DNAJA and DNAJB in particular have been implicated in protein quality control after stress⁷. The two classes are thought to interact with the Hsp70 machine separately from each other, chaperoning different types of client⁷. But if, and how, a combination of proteins of different DNAJ classes might act in tandem had not previously been addressed.

Using preformed, heat-aggregated model proteins, Nillegoda *et al.* show that DNAJAs and DNAJBs accelerate protein disaggregation synergistically through a mechanism that is distinct from their classical role in protein folding. In a series of experiments, the authors demonstrated that, rather than acting sequentially, the different DNAJ classes act in parallel with one another, and together with Hsp70, to mediate disaggregation.

DNAJ proteins are known⁷ to interact with Hsp70 through their evolutionarily conserved J-domains, and with their targets through variable carboxy-terminal domains (CTDs). Intriguingly, the authors report that the synergistic relationship between DNAJA and DNAJB during disaggregation depends on interactions between the J-domain of one protein and the CTD of the other (Fig. 1). This interaction is independent of the motif called HPD, through which DNAJ interacts with Hsp70, but is instead mediated by conserved, differently charged regions in the J-domain and CTD in each protein. Most DNAJ proteins studied so far act as homodimers (pairs of the same protein)⁷, and Nillegoda and colleagues propose that, in the disaggregation complex, a DNAJA homodimer binds a DNAJB

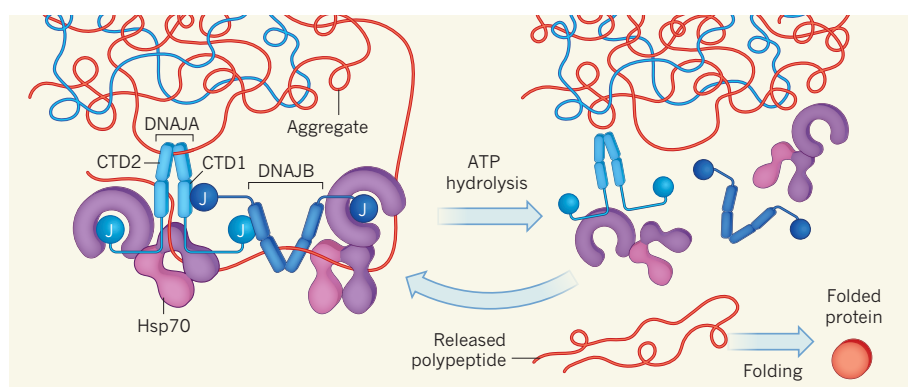


Figure 1 | Protein disaggregation by DNAJ proteins. Nillegoda *et al.*⁶ investigated the mechanism underlying protein disaggregation in animals. They report that two classes of DNAJ protein, DNAJA and DNAJB, acting as homodimers, use their carboxy-terminal domains (CTDs) to bind to aggregates. One molecule of the DNAJA homodimer binds to CTD1 of the adjacent DNAJB through a J-domain, and vice versa. The other subunit of each homodimer uses its J-domain to bind to one Hsp70 molecule, recruiting Hsp70 to bind to the aggregate. Entropic pulling forces, driven by Hsp70 binding to and hydrolysing ATP molecules, untangle a polypeptide molecule from the aggregate. The released polypeptide can then be folded into a normal protein. The DNAJ–Hsp70 complex breaks apart, and the process begins again.

homodimer. One protein from each homodimer engages in this interaction, and the other is free to interact with an Hsp70 molecule. Thus, two Hsp70 molecules can be recruited to each complex.

The authors suggest that the hydrolysis of ATP molecules by Hsp70 brings about entropic pulling, a process that untangles polypeptide molecules from the aggregate, allowing them to be refolded or degraded (Fig. 1). Finally, Nillegoda *et al.* demonstrate that the complex works most efficiently when the pre-formed aggregates have been generated in the presence of small Hsps. These proteins are excellent ATP-independent chaperones, and can join strings of protein monomers under conditions of acute stress, incorporating these unfolded proteins into an aggregate — a process that makes the aggregates themselves more accessible for disaggregation reactions⁸.

A basic question that remains to be resolved is why animals lost Hsp104 and evolved to use DNAJs for disaggregation instead. One speculation is that Hsp104 is rather promiscuous — it will efficiently break down any aggregated protein complex. In animal cells, this would include complexes that are necessary for cellular function, such as RNA granules, which are assembled by controlled aggregation and are needed to transport RNA along neurons⁹. By contrast, the complex discovered in this study seems to allow controlled, substrate-specific disaggregation.

The existence of an animal disaggregating complex provides an explanation for previously observed protein-disaggregation activity in mammals⁴. Furthermore, the complex acts in a similar way to the specialized DNAJ–Hsp70 combinations that disassemble normal, non-pathological protein complexes (for example, interactions between Hsp70 and the DNAJ auxilin mediate the removal of clathrin protein from certain vesicles¹⁰). Given the large number of DNAJ proteins

in animals, the authors' findings suggest that many combinations of DNAJs might interact, to make complexes that untangle specific clients.

Nillegoda and colleagues' data are restricted to tests on heat-induced protein aggregates, but it will be important to examine if and how DNAJ–Hsp70 complexes act on the various types of aggregate that cause human degenerative disorders⁷. In such diseases, only chaperones that inhibit the formation of aggregates have been described so far¹¹. Inhibition of aggregation might delay the onset of disease, but would require a continual treatment that

begins before the disease arises — an unfeasible strategy for diseases of spontaneous or unknown origin. By contrast, targeting disaggregation activity after the first symptoms of disease become apparent could be of therapeutic value. It will be necessary to determine the stage *in vivo* at which the DNAJ–Hsp70 complex acts to untangle aggregates brought about by different disease-associated proteins; to define the different DNAJ combinations involved; and to test whether the complex can alleviate aggregation-associated toxicity and thus halt disease progression. ■

Harm H. Kampinga is in the Department of Cell Biology, University Medical Center Groningen and University of Groningen, 9713 AV Groningen, the Netherlands. e-mail: h.h.kampinga@umcg.nl

1. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. *Nature* **475**, 324–332 (2011).
2. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. *Science* **319**, 916–919 (2008).
3. Doyle, S. M., Genest, O. & Wickner, S. *Nature Rev. Mol. Cell Biol.* **14**, 617–629 (2013).
4. Kampinga, H. H. *J. Cell Sci.* **104**, 11–17 (1993).
5. Rampelt, H. *et al.* *EMBO J.* **31**, 4221–4235 (2012).
6. Nillegoda, N. *et al.* *Nature* **524**, 247–251 (2015).
7. Kampinga, H. H. & Craig, E. A. *Nature Rev. Mol. Cell Biol.* **11**, 579–592 (2010).
8. Haslbeck, M., Franzmann, T., Weinfurter, D. & Buchner, J. *Nature Struct. Mol. Biol.* **12**, 842–846 (2005).
9. Wolozin, B. *Mol. Neurodegen.* **7**, 56–68 (2012).
10. McMahon, H. T. & Boucrot, E. *Nature Rev. Mol. Cell Biol.* **12**, 517–533 (2011).
11. Kakkar, V., Prins, L. C. & Kampinga, H. H. *Curr. Top. Med. Chem.* **12**, 2479–2490 (2012).

This article was published online on 5 August 2015.

LONGEVITY

Mapping the path to a longer life

Inhibiting the PI3K branch of the cell signalling induced by insulin and insulin-like growth factor can extend lifespan. The finding that inhibiting the RAS branch also extends lifespan in flies suggests a new target for anti-ageing drugs.

MORRIS F. WHITE

Understanding how cellular nutrient-sensing and homeostasis affect an organism's lifespan and susceptibility to cancer and degenerative diseases is clinically important but scientifically difficult. Work in model organisms, first in nematode worms¹ and then in fruit flies², has established that increasing the activity of the transcription factor FOXO, by reducing insulin and insulin-like growth factor signalling (IIS), can protect against cellular damage and ageing. Writing in *Cell*, Slack *et al.*³ now show that selective

mutation of the fruit fly IIS adaptor protein Chico to disrupt either of the two major IIS pathways — the RAS–ERK or the PI3K–AKT cascades — can extend the flies' lifespan. They find that inhibiting ERK signalling activates the transcriptional-repressor protein AOP, which, surprisingly, extends lifespan by just as much as the activation of FOXO that occurs during PI3K inhibition. This finding may have uncovered a new target for extending lifespan that lacks the adverse effects associated with PI3K inhibition, which include dysregulated metabolism, reduced growth and infertility.

From a clinical perspective, a reduced IIS

pathway that extends lifespan is different from the resistance to the hormone insulin that leads to metabolic syndrome — the medical term for a combination of diabetes, high blood pressure and obesity. The most frequent misconception is that increasing the amount of circulating insulin to overcome insulin resistance is a healthy goal. Although increasing circulating insulin levels can prevent high blood glucose levels and slow the progression to type 2 diabetes, it is associated with obesity and abnormal lipid levels and cardiovascular disease. Moreover, at the cellular level, chronic high insulin or long-term insulin therapy can retard tissue repair and maintenance by inhibiting autophagy, the process that removes damaged intracellular proteins. How to safely achieve a longer lifespan by reducing IIS in the face of high-sugar diets that promote insulin resistance and increase insulin levels is a challenging conundrum.

The fruit fly *Drosophila melanogaster* is useful for investigating the relationship between the IIS pathway and lifespan because it has a short life cycle and its genome encodes single copies of many IIS components. However, translating discoveries in fruit flies to other organisms can be tricky, because in people and other animals the IIS pathway is distributed between two homologous receptors (InsR and IGF1R), three or four adaptor proteins (IRS1, IRS2, IRS3 in rodents, and IRS4) and several effector proteins of the MAPK, PI3K, AKT and FOXO families. Some variants in the genes encoding IGF1R⁴ and FOXO3A⁵ have been associated with human longevity; however, complete loss of InsR or IGF1R function is fatal after birth for mice and people. Yet inactivation of InsR in murine adipose tissue extends lifespan⁶, as does deletion of IGF1R in the brain⁷. Thus the window for increasing lifespan by modulating IIS in mammals seems to be tissue specific.

The natural loss of growth-hormone receptors in humans with Laron syndrome (a type of dwarfism) causes obesity and reduces circulating insulin and IGF1, as well as decreasing the incidence of diabetes or cancer⁸. Notably, mice without growth-hormone receptors share similar traits and are the longest-lived laboratory mouse strain. Clearly, it is essential to understand the tissue-specific effects of reduced IIS in mammals to modulate lifespan with the fewest possible adverse effects.

A major problem with reduced IIS is the risk of dysregulated metabolism and growth associated with inhibiting the PI3K cascade⁹. Despite this, efforts to target the PI3K branch of IIS in ageing might be successful with a better understanding of which protein isoforms to target and with improved inhibitors. In the meantime, Slack *et al.* find that exposing flies to the small molecule trametinib, currently used for cancer therapy, achieves similar lifespan extension to the inhibition of the PI3K

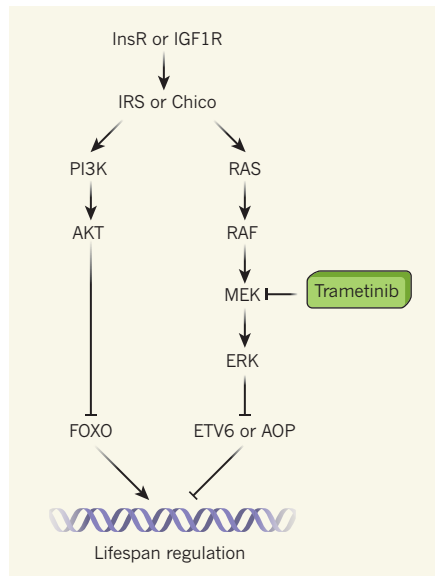


Figure 1 | IIS and lifespan. Insulin and insulin-like growth factor signalling (IIS) is triggered by binding of the insulin receptor (InsR) or insulin-like growth factor 1 receptor (IGF1R) and activation of an adaptor protein, which can be one of three or four IRS proteins in mammals or the Chico protein in *Drosophila* fruit flies. The pathway then splits into two main branches. One activates the proteins PI3K and AKT to cause inhibition of the transcription factor FOXO, an activator of gene transcription. The other branch activates the protein RAS, which then triggers a cascade involving RAF, MEK and ERK to inhibit the transcription factor ETV6 (or its equivalent in flies, AOP), a repressor of gene transcription. Slack *et al.*³ show that, in *Drosophila*, mutating *chico* to disrupt either of these pathways, or inhibiting the RAS–ERK pathway by treating flies with the MEK inhibitor trametinib, leads to extended lifespan. Adapted from ref. 3 (CC BY 4.0).

pathway. Trametinib inhibits the ERK branch of the IIS pathway by inhibiting the protein kinase enzyme MEK (Fig. 1).

The authors also show that the extension of fly lifespan by ERK or PI3K inhibition is not additive, which suggests that the two branches of the pathway might converge on modulating the expression of common genes that regulate lifespan¹⁰. Inhibition of ERK activates AOP, whereas inhibition of PI3K activates FOXO; both transcription factors do indeed bind a common subset of genes, but the exact targets that control lifespan are unknown¹⁰ (Fig. 1). Moreover, FOXO is usually a transcriptional activator, whereas AOP is a repressor that opposes the activity of another factor, PNT, in *Drosophila*. Interestingly, coactivation of FOXO and PNT can have detrimental effects that are attenuated by AOP¹⁰, indicating that favourable crosstalk between AOP and FOXO might modulate common genes needed to extend lifespan.

Despite the potential to bypass PI3K inhibition, it remains to be investigated whether inhibiting the ERK cascade can extend mammalian lifespan without any adverse effects. ERK is a member of the MAPK enzyme

family, which mediates cellular responses to a wide range of extracellular cues to regulate cell growth, differentiation and survival. Although MEK inhibition has been shown¹¹ to improve glucose tolerance in diet-induced obese mice, the tissues in which this is beneficial are ill defined. Further work is needed to establish whether inhibiting the ERK signalling branch is a plausible mechanism-based strategy for extending lifespan, especially when started in adults.

While work continues to devise medical strategies to extend lifespan, calorie restriction remains the best-known way to increase lifespan in yeast, nematodes, fruit flies, rodents and some primates¹². Calorie restriction can reduce the progression of age-related diseases, including obesity, insulin resistance, type 2 diabetes, cardiovascular disease and cancer, but it is difficult for people to use in the long term and can be dangerous if unmonitored or used to excess; furthermore, its beneficial effects on human lifespan are unproven¹³. Recent work suggests that a fasting-mimicking diet that produces intermittent, brief bouts of calorie restriction can produce health benefits in people and extend the lifespan of mice¹⁴.

Both these approaches increase insulin sensitivity and reduce circulating insulin and IGF1 concentrations, and so reduced IIS might be involved in the observed effects. Whether medical strategies — such as using the drug rapamycin to inhibit the enzyme TOR¹⁵, the antidiabetic drug acarbose to mimic calorie restriction¹⁶, or inhibiting the ERK cascade as described by Slack and colleagues — can exploit or augment the underlying molecular mechanisms engaged by calorie restriction and reduced IIS in the face of nutrient excess is an important area for future investigation. ■

Morris F. White is in the Division of Endocrinology, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.
e-mail: morris.white@childrens.harvard.edu

1. Kenyon, C. *Phil. Trans. R. Soc. B* **366**, 9–16 (2011).
2. Hwangbo, D. S. *et al. Nature* **429**, 562–566 (2004).
3. Slack, C. *et al. Cell* **162**, 72–83 (2015).
4. Suh, Y. *et al. Proc. Natl Acad. Sci. USA* **105**, 3438–3442 (2008).
5. Willcox, B. J. *et al. Proc. Natl Acad. Sci. USA* **105**, 13987–13992 (2008).
6. Blüher, M., Kahn, B. B. & Kahn, C. R. *Science* **299**, 572–574 (2003).
7. Kappeler, L. *et al. PLoS Biol.* **6**, e254 (2008).
8. Guevara-Aguirre, J. *et al. Science Transl. Med.* **3**, 70ra13 (2011).
9. Thorpe, L. M., Yuzugullu, H. & Zhao, J. J. *Nature Rev. Cancer* **15**, 7–24 (2015).
10. Alic, N. *et al. PLoS Genet.* **10**, e1004619 (2014).
11. Banks, A. S. *et al. Nature* **517**, 391–395 (2015).
12. Anisimov, V. N. & Bartke, A. *Crit. Rev. Oncol. Hematol.* **87**, 201–223 (2013).
13. Fontana, L., Partridge, L. & Longo, V. D. *Science* **328**, 321–326 (2010).
14. Brandhorst, S. *et al. Cell Metab.* **22**, 86–99 (2015).
15. Harrison, D. E. *et al. Nature* **460**, 392–395 (2009).
16. Harrison, D. E. *et al. Aging Cell* **13**, 273–282 (2014).

Universal allosteric mechanism for $G\alpha$ activation by GPCRs

Tilman Flock¹, Charles N. J. Ravarani^{1*}, Dawei Sun^{2,3*}, A. J. Venkatakrishnan^{1†}, Melis Kayikci¹, Christopher G. Tate¹, Dmitry B. Veprintsev^{2,3} & M. Madan Babu¹

G protein-coupled receptors (GPCRs) allosterically activate heterotrimeric G proteins and trigger GDP release. Given that there are ~800 human GPCRs and 16 different $G\alpha$ genes, this raises the question of whether a universal allosteric mechanism governs $G\alpha$ activation. Here we show that different GPCRs interact with and activate $G\alpha$ proteins through a highly conserved mechanism. Comparison of $G\alpha$ with the small G protein Ras reveals how the evolution of short segments that undergo disorder-to-order transitions can decouple regions important for allosteric activation from receptor binding specificity. This might explain how the GPCR- $G\alpha$ system diversified rapidly, while conserving the allosteric activation mechanism.

G proteins bind guanine nucleotides and act as molecular switches in a number of signalling pathways by interconverting between a GDP-bound inactive and a GTP-bound active state^{1,2}. They consist of two major classes: monomeric small G proteins³ and heterotrimeric G proteins⁴. While small G proteins and the α -subunit ($G\alpha$) of heterotrimeric G proteins both contain a GTPase domain (G-domain), $G\alpha$ contains an additional helical domain (H-domain) and also forms a complex with the $G\beta$ and $G\gamma$ subunits. Although they undergo a similar signalling cycle (Fig. 1), their activation differs in one important aspect. The guanine nucleotide exchange factors (GEFs) of small G proteins are largely cytosolic proteins, whereas the GEFs of $G\alpha$ proteins are usually membrane-bound GPCRs. While GEFs of small G proteins interact directly with the GDP binding region^{1,3}, GPCRs bind to $G\alpha$ at a site

almost 30 Å away from the GDP binding region⁵ and allosterically trigger GDP release to activate them.

The high-resolution structure of the $G\alpha_s$ -bound β_2 -adrenergic receptor (β_2 AR)⁵ provided crucial insights into the receptor-G protein interface and conformational changes in $G\alpha$ upon receptor binding^{6,7}. Recent studies described dynamic regions in $G\alpha_s$ ⁶ and $G\alpha_i$ ⁸, the importance of displacement of helix 5 (H5) of $G\alpha_s$ and $G\alpha_t$ by up to 6 Å into the receptor⁵, the extent of helical domain opening during GDP release^{9,10}, and identified residues that contribute to $G\alpha_i$ activation^{7,10}. These studies focused on single, specific $G\alpha$ proteins; however, in humans there are 16 different $G\alpha$ genes, with at least 21 isoforms⁴ that can be grouped into four functional subfamilies ($G\alpha_s$, $G\alpha_i$, $G\alpha_q$, $G\alpha_{12}$), which each regulate different signalling pathways¹¹. Although they belong to the same protein fold, they have diverged significantly in their sequence such that each $G\alpha$ protein can be specifically activated by one or several of the ~800 human GPCRs⁴. Thus, a fundamental question is whether there is a universal mechanism of allosteric activation that is conserved across all $G\alpha$ protein types¹⁰. Allosteric communication in proteins is mediated through conformational changes, which are facilitated by the re-organization of non-covalent contacts between residues. Thus, studying these contacts can provide detailed insights into the mechanism of allostery^{12–16}. On the basis of a comprehensive analysis, here we propose that GPCRs interact and activate $G\alpha$ subunits through a conserved mechanism. We describe molecular details of the key structural transitions and pinpoint residues that constitute the ‘common core’ of $G\alpha$ activation.

Common $G\alpha$ numbering and residue contact networks

We created a structural and sequence alignment of 80 $G\alpha$ structures from diverse organisms and 973 sequences from 66 species that have a GPCR-G protein system (auto-activating plant $G\alpha$ proteins were not considered; Methods). To enable the comparison of any residue/position between different $G\alpha$ proteins, we devised the common $G\alpha$ numbering (CGN) system (Fig. 2a). The CGN provides an ‘address’ for every residue in the DSP format, referring to: (1) the domain (D); (2) the consensus secondary structure (S); and (3) the position (P) within the secondary structure element. For instance, phenylalanine 336 in $G\alpha_{i1}$ is denoted as Phe336^{G.H5.8} as it is the eighth residue within the consensus helix H5 of the G-domain. The corresponding position in $G\alpha_{s2}$ is

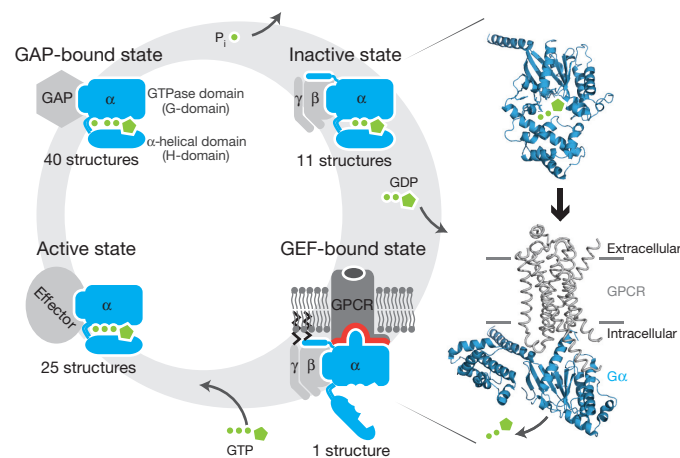


Figure 1 | $G\alpha$ signalling states and activation. Heterotrimeric G proteins (1) release GDP upon binding to a guanine nucleotide exchange factor (GEF), which are G-protein-coupled receptors, (2) bind GTP and recruit downstream effectors, and (3) hydrolyse GTP, promoted by a GTPase activating protein (GAP), leading to (4) the inactive, GDP-bound state. Structures of the $G\alpha$ subunit (blue) bound to GDP (Protein Data Bank (PDB) accession 1got; inactive state; top) and bound to the β_2 -AR (grey) (PDB 3sn6; active state; bottom) are shown.

¹MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK. ²Laboratory of Biomolecular Research, Paul Scherrer Institut, 5232 Villigen, Switzerland. ³Department of Biology, ETH Zurich, 8039 Zurich, Switzerland. [†]Present address: Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA.

*These authors contributed equally to this work.

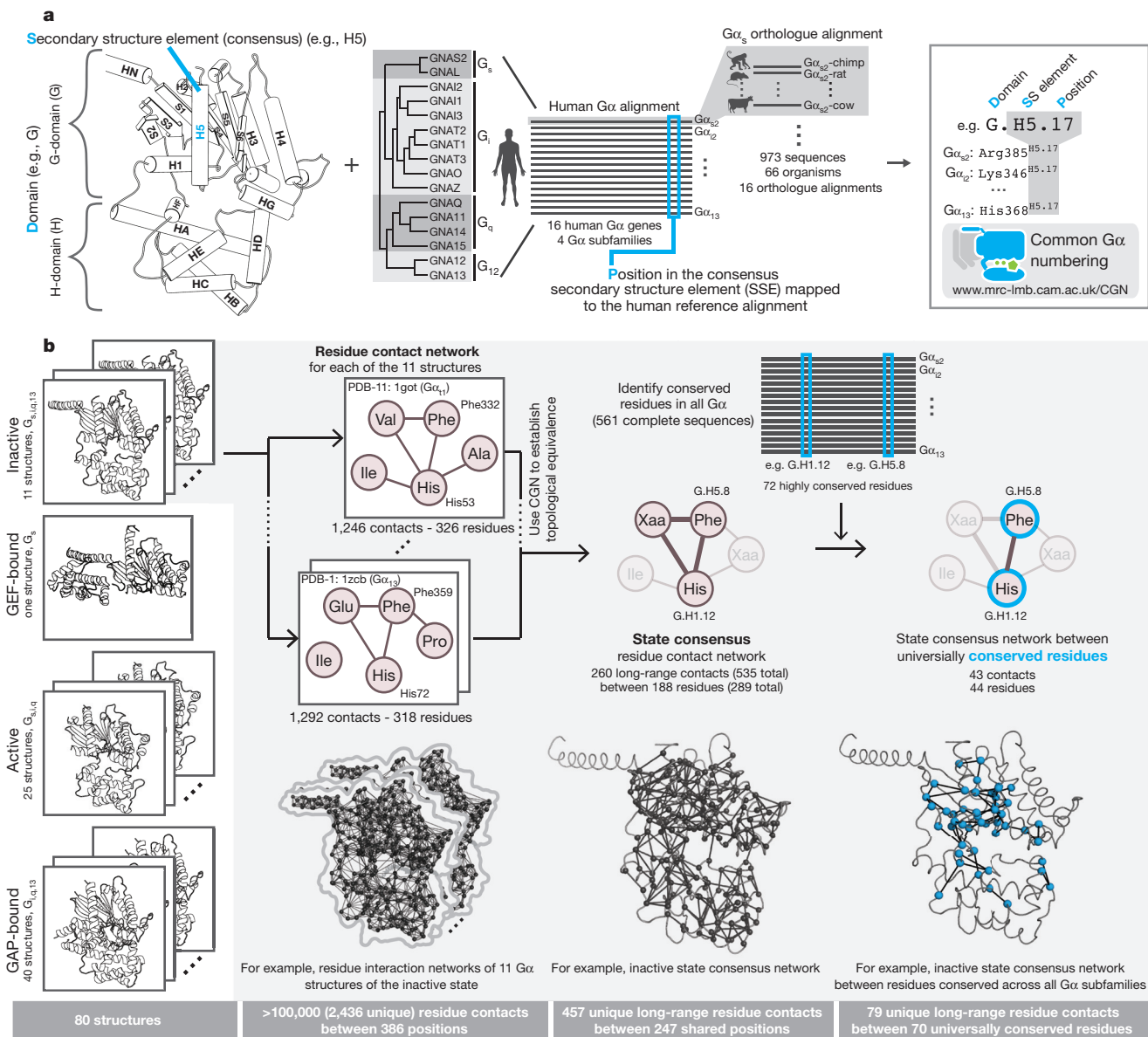


Figure 2 | The common G α numbering (CGN) system and G α conserved contact networks. **a**, Every position in a G α is denoted by its domain (D), the consensus secondary structure element where it is present (S), and position (P) within the consensus SSE. Names of SSEs are shown in the cartoon; loops are named with lowercase letters of the flanking SSEs (for example, h1ha; see Extended Data Fig. 1 for all SSEs). An alignment of all 973 G α proteins

Phe376^{G.H5.8}. Loops are labelled in lowercase letters of their flanking secondary structure elements (SSE); for example, s6h5 refers to the loop connecting strand S6 with helix H5 (see Extended Data Fig. 1, Methods and Supplementary Note). A CGN mapping webserver is available at <http://www.mrc-lmb.cam.ac.uk/CGN>.

The G α structures were assigned to the four major signalling states (Fig. 1) and the non-covalent contacts between residues were calculated for each structure. We computed the consensus non-covalent residue contacts from all G α structures of the same signalling state. Using the CGN, we integrated information on evolutionary conservation for every position and derived the consensus contacts mediated by universally conserved residues for each signalling state (Fig. 2b). We find that each step of the signalling cycle undergoes contact re-organization to variable extents. Since these conformational changes involve conserved residues, the observed contact re-organization is likely to be universal for all G α proteins. A description and additional interpretations are provided in

belonging to all the 4 subfamilies from 66 species allowed the identification of equivalent residues. An online webserver allows mapping of any G α sequence or structure to the CGN system (<http://www.mrc-lmb.cam.ac.uk/CGN>).

b, Computation of consensus residue contact networks between conserved residues for different G α signalling states. See Methods, Supplementary Note and Supplementary Data.

the Supplementary Note and Supplementary Information. Below, we describe the major findings pertaining to G α activation. We first focus on the GPCR–G α interface and then describe the molecular details of how the non-covalent contacts are re-organized and propagated to the GDP binding pocket, leading to GDP release.

GPCR–G α protein interface

Analysis of the buried surface area (BSA) and residue contacts between the β_2 AR–G α_s interface⁵ shows that H5 contributes $\sim 70\%$ (845 Å²; 15 residues) of the total BSA. Other SSEs (s2s3, h4hg, H4, h4s6, S6) cover $\sim 20\%$ (289 Å²; 14 residues), and the amino-terminal membrane-anchored helix HN and its loop with strand S1 contribute $\sim 10\%$ (120 Å²; 5 residues) of the total BSA (Fig. 3a). H5 is the key interface element that contacts residues in transmembrane helices (TM3, TM5 and TM6) and intracellular loops 2 and 3 (ICL2 and ICL3) of the β_2 AR⁵. Contacts from the other G α regions are mainly restricted to ICL3 of the

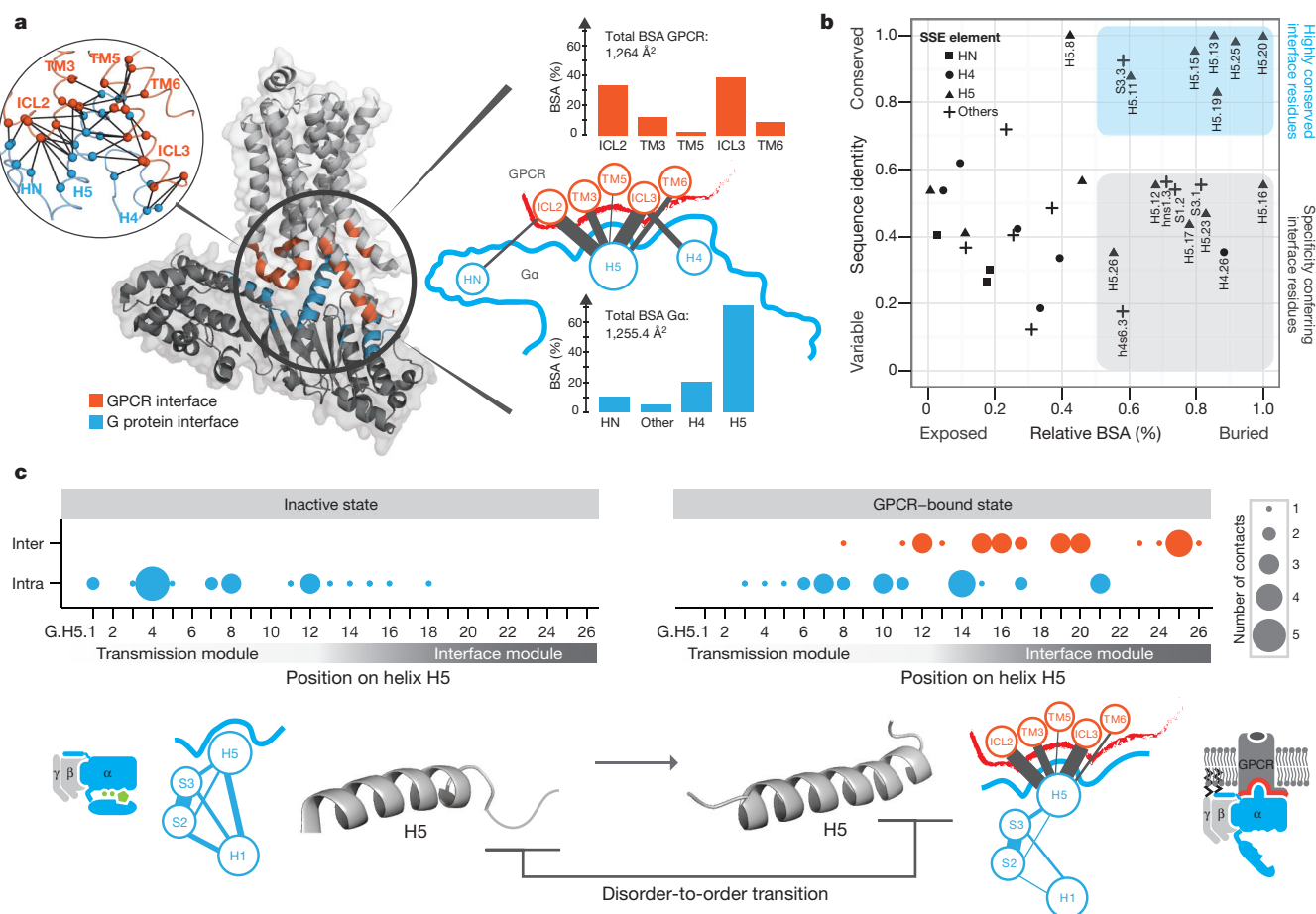


Figure 3 | Helix 5 contains the conserved interface region and is comprised of two modules. **a**, Inter $G\alpha$ -GPCR residue contact network (inset) and buried surface area (BSA) analysis of the heterotrimeric $G\alpha_s$ - β_2 AR structure (3sn6). The line width between the nodes (SSE) denotes the number of consensus residue contacts. GPCR positions are denoted by extending the Ballesteros-Weinstein numbering system (note parts of ICL3 become extended TM5 in the active state of the receptor). **b**, Scatter-plot of $G\alpha$ sequence conservation and

normalized BSA highlights the conserved and variable interface residues. **c**, Consensus contact rewiring between the inactive and the GPCR-bound state by H5 residues. Positions mediating intra $G\alpha$ protein contacts (blue) and receptor-mediated contacts (red) are shown. The circle size represents the number of contacts. H5 can be divided into transmission and interface module. The disorder-to-order transition of the H5 C-terminal region upon receptor binding and SSE contact rewiring are shown.

receptor. An analysis of the contacts of the $G\alpha_t$ carboxy-terminal peptides bound to rhodopsin^{17–20} shows that conserved residues in H5 tend to interact with the corresponding, topologically equivalent residues in rhodopsin (Supplementary Data).

We mapped evolutionary conservation onto the β_2 AR- $G\alpha_s$ interface and found that H5 is the only interface region that harbours residues that are highly conserved across species and $G\alpha$ protein types (~27% of H5; 7 residues; Fig. 3b). These residues have significantly higher BSA compared to the non-conserved H5 interface residues. Several highly conserved $G\alpha$ interface residues interact with conserved interface residues on the GPCR (Extended Data Fig. 2a). Computational energy calculations show that these residues make the highest interface energy contribution, suggesting that they are important for complex formation (Extended Data Fig. 2a). Thus, the universally conserved residues on H5 might form the conserved 'interaction hotspots' for different $G\alpha$ proteins to interact with their cognate receptors in a similar binding mode. We also found that two-thirds of the H5 residues are variable but half of these (8 residues) still contact the receptor. Thus, H5 harbours distinct sets of interface residues that are either conserved or variable across the different $G\alpha$ proteins. The variable positions on H5 together with the other interface regions (Fig. 3b) could be important for selective coupling to different receptors, as shown for individual $G\alpha$ proteins^{21,22}. This suggests that the conserved $G\alpha$ interface positions provide the basis for a common mode of receptor binding, while the variable positions might confer selectivity in receptor coupling (Supplementary Note).

The role of helix H5 in $G\alpha$ activation

In the 79 structures of $G\alpha$ not bound to a GPCR, the C-terminal residues of H5 are characterized by missing electron density (Extended Data Fig. 2b). This region undergoes a disorder-to-order transition and extends H5 upon receptor binding as shown for $G\alpha_s/t_l$ ^{5,6,8,20,23}. Analysis of H5 from 561 full-length $G\alpha$ homologues suggests that the higher disorder propensity of the last eight residues compared to the rest of H5 is a universal feature (Extended Data Fig. 2b). Within this disordered region, hydrophobic positions Ile^{H5.15}, Leu^{H5.20} and Leu^{H5.25} contact the GPCR, fold into a helix and are conserved between human proteins and the yeast homologue Gpa1 (~1,200 million years (Myr) ago). This highly conserved peptide motif²⁴ is involved in the disorder-to-order transition upon receptor binding and suggests that this structural transition mediated by the three key conserved interface residues is likely to be a universal feature of all G proteins.

To understand the effect of GPCR binding on contact re-organization within $G\alpha$, we analysed the consensus contacts in the inactive state (11 structures) and the active state (β_2 AR- $G\alpha_s$ structure). Although we had only one structure of the receptor- $G\alpha$ protein complex, identifying the differences between the contacts of the active state and the consensus contacts of the inactive state allowed us to focus on residues that are re-organized upon receptor binding (and hence likely to be relevant for all $G\alpha$ types; Supplementary Note). In all inactive state structures, the N-terminal part of H5 makes extensive contacts with a number of SSEs within the G-domain of $G\alpha$. Two universally conserved positions

(Phe^{H5.8} and Val^{H5.7}) on H5 contact conserved positions in H1, S2 and S3 (left lobe), and S5 and S6 (right lobe), respectively (Extended Data Fig. 3). In the GPCR-bound state, H5 loses 20% of its intra-G α contacts (primarily with H1), and gains 27 intermolecular residue contacts with the GPCR (Fig. 3c). Upon receptor binding, the H5 contacts with the right lobe are not lost, but are re-organized to accommodate the structural changes and might be important for the stability of the receptor-bound complex (for discussion, see Supplementary Note and ref. 25). Thus, H5 is composed of two highly conserved modules with distinct functions; that is, an interface module important for receptor binding, and a transmission module that harbours intra-G-protein contacts, which are re-organized upon receptor binding (Fig. 3c).

In contrast to H5, the non-conserved interface regions from H4, h4s6 and h4hg undergo less marked re-organization of intra-G α contacts upon receptor binding (Extended Data Fig. 3). This suggests that the conserved mechanism of allosteric activation is primarily mediated by the movement of H5, thereby breaking the contacts between H5 and H1. As the residues that form these contacts are conserved in all 16 G α proteins, the described contact re-organization is likely to be universal for all G α proteins.

The role of helix H1 in G α activation

In the inactive state, H1 acts as a structural ‘hub’ by linking different functional regions of G α . H1 contacts the N-terminal part of H5 (transmission module), H-domain and GDP through universally conserved residues (Fig. 4a). In this manner, H1 links the H-domain and the GDP binding site with the conserved residues in the H5 transmission module, which in turn is physically linked to the H5 interface module that binds the receptor. The conserved consensus contacts between H5 and H1 seem essential for the structural integrity of H1. Computational calculations of the per-residue contribution to protein stability for the 79 non-receptor-bound structures are consistent with their role in stabilizing H1 (Extended Data Fig. 4). Upon GPCR binding, the H5–H1 contacts are lost, affecting the structural integrity of H1. The contact-mediating positions in H1 have missing electron density in the β_2 AR–G α structure⁵ and hydrogen/deuterium exchange experiments have shown that this region is dynamic in G α_s upon receptor binding⁶. Upon becoming

flexible, the conserved consensus contacts between H1, GDP and the H-domain hinge region are lost; this results in the loss of a significant fraction of all contacts made with GDP, thereby weakening its binding affinity (Fig. 4b), and results in increasing the likelihood of H-domain opening. Since the entire sequence of H1, the contacting residues on H5, and the H-domain hinge positions are highly conserved across all G α proteins, the mechanism of GDP release is likely to be universal for all G α proteins.

Universal mechanism of G α activation

While variable interface residues in H5 and elsewhere allow specific binding to distinct GPCRs, we find that H5 primarily harbours conserved positions that might allow a common mode of receptor binding and a conserved mechanism of allosteric activation. The contact re-organization between conserved residues links the disorder-to-order transition of H5 upon receptor binding to a change in structural stability of H1, ultimately leading to GDP release. More specifically, H5 is divided into: (1) The N-terminal transmission module, which forms a π – π cluster linking H5, S2, S3 and H1 via universally conserved residues Phe^{H5.8}, His^{H1.12}, Phe^{S2.6} and Phe^{S3.3} in the inactive state; and (2) the C-terminal interface module, which undergoes a disorder-to-order transition in the intracellular cavity of the receptor via universally conserved positions. This structural transition results in a displacement of the H5 transmission module, thereby interrupting the π – π cluster. The re-organized residues in the cluster (Phe^{H5.8} and Phe^{S3.3}) contact conserved residues within ICL2 of the receptor (extrapolated Ballesteros–Weinstein numbering: 3.58 and 3.57 of β_2 AR⁵), as confirmed recently for the CB2 receptor–G α_i complex²⁶. Since the conserved π – π cluster is important for the structural integrity of H1, its disruption leads to an increased flexibility of H1. H1 has a central role in the inactive state by forming contacts both to GDP (Extended Data Fig. 4) via the Walker A motif^{2,27} and to the H-domain (through a ‘cation– π hinge’ motif; Extended Data Fig. 5). Thus, the increased flexibility due to the partial unfolding of H1 facilitates GDP release and H-domain opening. The only other conserved inter-domain contact is an ‘ionic latch’ between the C-terminal loop of helix HG of the G-domain and the hChD loop of the H-domain. This contact is broken upon receptor binding, which might be a result of the reorganization of the right G α lobe (Extended Data Fig. 3).

In addition to H1, residues around HG and within the s6h5 loop (TCAT motif) contact the GDP (Extended Data Fig. 3). The conserved guanine-contacting TCAT motif preserves many of its contacts within G α upon receptor binding, although TCAT-to-H1 contacts are lost and new contacts are formed between H5, S5, S6 and the TCAT motif (that is, the re-organized right lobe). Likewise, residues that contact the guanosine moiety, including the interaction between the TCAT motif and HG, s1h1 (P-loop), S5 and S6, are not extensively re-organized during G α activation. Whether this arrangement poises G α for GTP binding (which differs from GDP by a single phosphate and whose physiological concentration exceeds GDP several-fold²²), and whether GTP has the capacity to stabilize H1 on its own and trigger G α release from the receptor, remains to be addressed. An analysis of the GTP-bound G α reveals that the presence of the third phosphate facilitates additional contacts with the switch regions (Extended Data Fig. 4c).

Conceptually, the GPCR-bound G α conformation can be considered as a metastable G α transition state that is stable only due to interactions with the GPCR. Thus, the lost intra-G α contacts between H1 and the transmission module of H5 are compensated by the helix extension of H5, the gained receptor interface contacts, and some re-organized contacts in the right lobe of G α (H5 with S5, S6; see role of conserved Y320^{G.S6.2} in ref. 25). In this manner, H5 and H1 act as the primary conduits of information transfer between the receptor interaction interface (input) and the GDP binding site (output). The residues of the structural motifs and functional elements described here are conserved in all the different families of G α proteins (Extended Data Fig. 5a). Thus, the mechanism described here is likely to be universal for activation of all cognate GPCR and G α protein pairs.

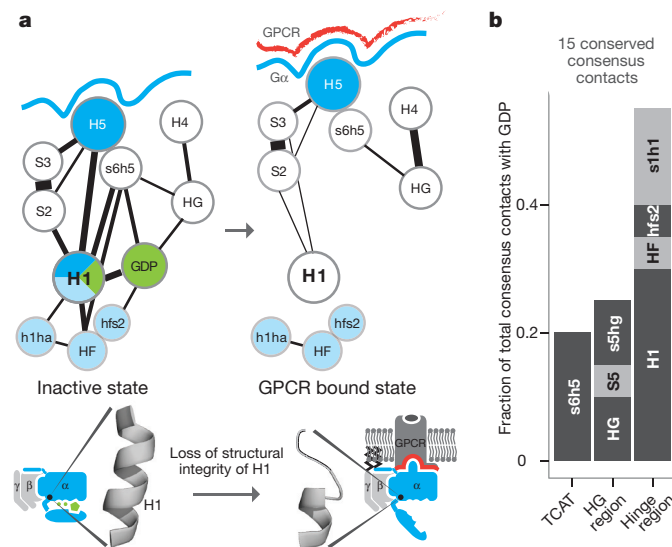


Figure 4 | Helix H1 is the key SSE that contacts H5, GDP and the H-domain. **a**, Consensus SSE contacts involving H1. The line width between the nodes (SSE) denotes the number of consensus residue contacts. Upon receptor binding, H5 is displaced and crucial contacts between H1 and H5 (dark blue) are lost. This might explain the increased flexibility of H1 in the GPCR-bound state, which results in the loss of GDP contacts (green) and the H-domain hinge region (formed by H1, h1ha and HF; light blue). **b**, The extent of GDP consensus contacts mediated by the different SSEs. See Extended Data Fig. 4.

Disease and engineered $G\alpha$ mutations

We analysed disease-causing mutations in the human population and found three key positions that are mutated (Supplementary Table 3), resulting in constitutive $G\alpha$ activity: (1) a variant of the transmission module residue Phe^{H5.8} in $G\alpha_{11}$ causes autosomal dominant hypocalcaemia type 2 by becoming constitutively active²⁸, possibly by destabilizing the contact between H5 and H1; (2) $G\alpha_i$ variant Ala^{S6H5.3}Ser, a position important for H1 stabilization and GDP binding, causes testotoxicosis by constitutively activating adenylate cyclase²⁹; and (3) $G\alpha_{11}$ variant Arg^{H1.9}Cys causes autosomal dominant hypo-parathyroidism³⁰, possibly by affecting H-domain opening and GDP release. We also analysed previously performed perturbation experiments on different $G\alpha$ types using the CGN system and can explain how these mutations might affect the activation mechanism (Fig. 5a and Supplementary Note). Furthermore, comprehensive alanine-scanning mutagenesis performed on $G\alpha_{11}$, coupled with thermostability assays of the mutants in the GDP-bound or nucleotide-free state coupled to rhodopsin²⁵, is also consistent with the analysis performed here (Fig. 5b). For example, mutations in the H5 interface module mainly affect $G\alpha_{11}$ –rhodopsin complex stability, whereas mutations in the H5 transmission module primarily affect the nucleotide-bound state of $G\alpha_{11}$. Alanine mutations in H1 highly destabilize the GDP-bound state, but not the $G\alpha_{11}$ –rhodopsin complex. In addition, mutating residues in the conserved π – π cluster that interact with Phe^{H5.8}

significantly affects $G\alpha_{11}$ stability in the GDP-bound form, but not in the receptor-bound form (see ref. 25). Taken together, our analysis, the CGN numbering system and the universal activation mechanism described here provides a unified framework to relate and interpret a number of independent experimental and disease mutations in different $G\alpha$ subfamilies.

Evolution of $G\alpha$ activation mechanism

To understand how allosteric regulation might have evolved in $G\alpha$, we compared the crystal structures of each equivalent signalling state of $G\alpha$ and Ras. We found that H5 and H1 have significantly changed their functional role. In Ras, H5 has a disordered extension (hyper-variable region) that is post-translationally modified for membrane anchoring³¹. In $G\alpha$, the equivalent region forms the GPCR interface module (the N-terminal disordered region of HN is the membrane anchor³²). The nucleotide-binding N-terminal part of H1 is conserved in its sequence and its structural orientation in both Ras and $G\alpha$ (Fig. 6 and Extended Data Fig. 6). In contrast, the central part of H1, which contacts the H-domain hinge, is only conserved in $G\alpha$. The C-terminal part of H1 has a different role in Ras and $G\alpha$. While the last turn of H1 in Ras folds back to bind the guanine moiety via a conserved π – π stack, the equivalent region in $G\alpha$ forms the metastable part of H1 that remains helical in the GDP-bound inactive state due to contacts with the N-terminal transmission module of H5. These contacts are missing in Ras, as the C terminus of H1 and the N terminus of H5 are each three residues shorter, and H1 in Ras is stable without the interactions with H5. This means that although Ras and $G\alpha$ are evolutionarily related and share the same architecture, minor but crucial differences in the number and pattern of non-covalent contacts between H5 and H1 have allowed the emergence of an allosteric mechanism for GDP release in $G\alpha$. Small extensions in H1 and H5 permit H1 to sense whether H5 is bound to the GPCR. The disordered C-terminal tail of H5 provides both conserved and variable interface residues that allow for a conserved $G\alpha$ activation mechanism and yet permit the evolution of receptor-binding specificity.

Discussion

Our analysis suggests that GPCRs interact and activate $G\alpha$ subunits through a highly conserved mechanism in which the interruption of the contacts between H1 and H5 is a key step for GDP release. In this sense, while H1 is the molecular switch for GDP release, H5 is the distal trigger that is ‘pulled’ upon receptor binding. Given that $G\alpha$ proteins belong to the same fold, the existence of evolutionarily conserved residues per se is not surprising. However, the observation that (1) the conserved residues form a network of non-covalent contacts that links the GPCR-binding site with the GDP-binding pocket and that (2) this network of contacts is consistently re-organized upon receptor binding suggests that this mechanism might constitute the common conserved set of structural changes for the allosteric release of GDP (Supplementary Note). While the conserved residue contacts are crucial for $G\alpha$ activation, non-conserved positions can still be important for allosteric activation in distinct $G\alpha$ proteins¹⁰. Thus, the identified residues are necessary but not sufficient for G-protein activation. The variable interface residues, as well as the $\beta\gamma$ subunits, could have important roles in receptor binding specificity for individual proteins. Thus, the conserved universal mechanism probably represents the ‘skeleton’ that can be incorporated into different contexts in different $G\alpha$ proteins to maintain a conserved mechanism of allosteric activation and yet permit specific binding to the receptor.

A comparison to small G proteins revealed how $G\alpha$ evolved to bind GPCRs at a site that is distal to the GDP binding pocket. Emergence of short regions in H5 and H1 that can undergo structural transitions seem to have been co-opted to make a new GEF (receptor) interface and link it to the GDP binding site. In such a system, displacement of a secondary structure element (H5) upon receptor binding can transmit information by re-organizing key non-covalent contacts

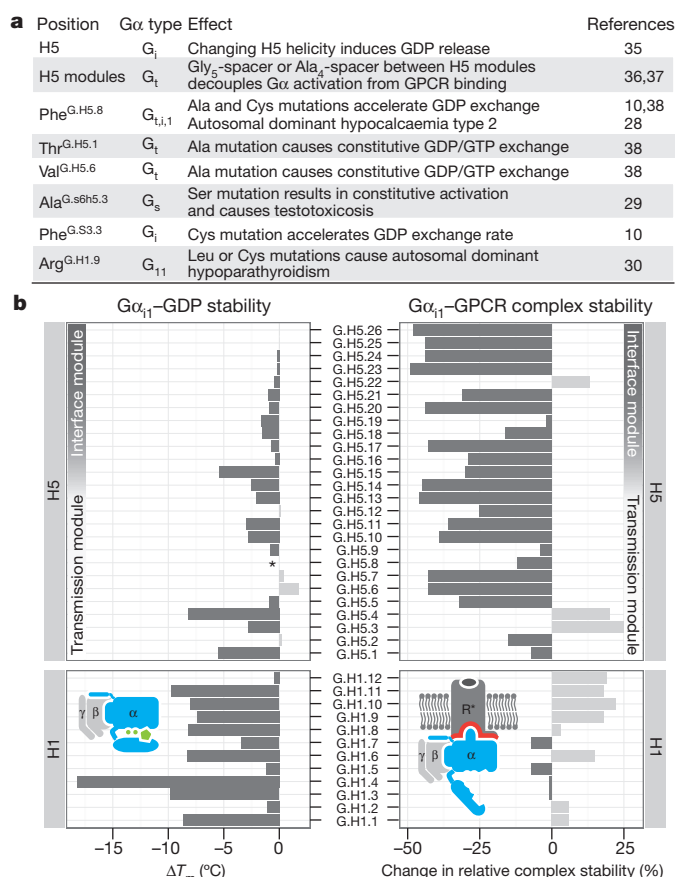


Figure 5 | Mutational studies support the universal $G\alpha$ activation mechanism. **a**, Disease and engineered mutations can be explained by the model of $G\alpha$ activation in different $G\alpha$ subfamilies. References^{10, 28–30, 35–38} are cited in this figure. **b**, Comparison of the stability of $G\alpha$ –GDP (ΔT_m ; °C) and the $G\alpha\beta\gamma$ –GPCR complex (change in relative complex stability; %) by $G\alpha_{11}$ alanine mutagenesis of every position in H1 and H5. Asterisk indicates that mutant Phe^{H5.8}A is not stable in the receptor-free state but can still form the complex with the receptor.

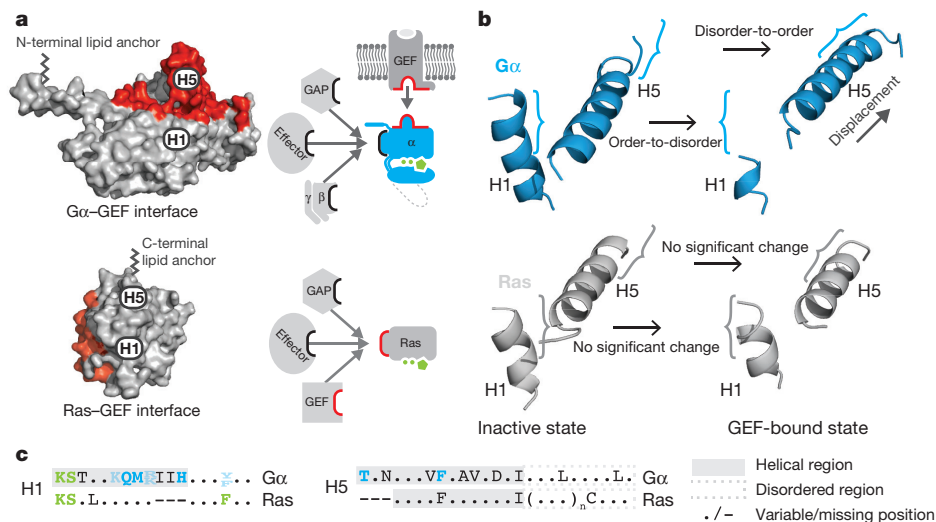


Figure 6 | H5–H1 interaction permits the allosteric activation mechanism. **a**, GEF interaction surfaces (red) for Gα (3sn6) and the small G protein human HRas (PDB 1bkd) in the same orientation. **b**, H1 and H5 of the inactive (1got, 4q21) and GEF-bound state (3sn6, 1bkd) for Gα (blue) and HRas (grey). **c**, Consensus sequences of equivalent residues of H5 and H1 in Gα and Ras. The

helical region is highlighted in a grey background and the H5 disordered region is shown in a dashed grey box. Conserved residues that contact GDP (green), the H-domain (light blue) or form contacts between H5 and H1 (dark blue) are highlighted. See Extended Data Fig. 6.

between conserved residues that connect different secondary structure elements. Thus, a common ancestor of the GTPase fold might have provided the structural framework that can be perturbed by GPCR binding through the interruption of H5–H1 contacts. This ‘new’ allosteric site for GEFs is physically separated from the Gα effector and regulator binding interfaces (Fig. 6a), and could have provided the basis for the expansion of the GPCR family without affecting the downstream signalling factors.

Our findings suggest that in addition to evolving extensive interfaces and allosteric ‘wires’ as observed in other proteins^{13,33}, another solution for allosteric communication is evolving short segments that undergo disorder-to-order transitions upon a trigger (for example, receptor binding). This mechanism involves the re-organization of a network of existing contacts to induce conformational changes that affect a distal site without the requirement of directly contacting residues but linked by the same secondary structure (for example, the interface and transmission module of H5 make distinct sets of contacts but are linked through the protein backbone; like a puppet on a string). Since disordered and loop regions tolerate more sequence changes than structured regions³⁴, an important implication is that such segments allow for the independent evolution of regions that are important for binding specificity but still maintain a conserved allosteric activation mechanism. Generalizing this principle, we suggest that disordered segments that can undergo structural transitions (regulated folding or unfolding) and thereby re-organize existing networks of contacts within structured regions of proteins could have an important role in other protein families and may be exploited in protein engineering applications.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 November 2014; accepted 16 June 2015.

Published online 6 July 2015.

- Vetter, I. R. & Wittinghofer, A. The guanine nucleotide-binding switch in three dimensions. *Science* **294**, 1299–1304 (2001).
- Leipe, D. D., Wolf, Y. I., Koonin, E. V. & Aravind, L. Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* **317**, 41–72 (2002).
- Rojas, A. M., Fuentes, G., Rausell, A. & Valencia, A. The Ras protein superfamily: evolutionary tree and role of conserved amino acids. *J. Cell Biol.* **196**, 189–201 (2012).

- Anantharaman, V., Abhiman, S., de Souza, R. F. & Aravind, L. Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system. *Gene* **475**, 63–78 (2011).
- Rasmussen, S. G. et al. Crystal structure of the β₂ adrenergic receptor–Gs protein complex. *Nature* **477**, 549–555 (2011).
- Chung, K. Y. et al. Conformational changes in the G protein Gs induced by the β₂ adrenergic receptor. *Nature* **477**, 611–615 (2011).
- Preininger, A. M., Meiler, J. & Hamm, H. E. Conformational flexibility and structural dynamics in GPCR-mediated G protein activation: a perspective. *J. Mol. Biol.* **425**, 2288–2298 (2013).
- Oldham, W. M., Van Eps, N., Preininger, A. M., Hubbell, W. L. & Hamm, H. E. Mechanism of the receptor-catalyzed activation of heterotrimeric G proteins. *Nature Struct. Mol. Biol.* **13**, 772–777 (2006).
- Westfield, G. H. et al. Structural flexibility of the Gαs α-helical domain in the β₂-adrenergic receptor Gs complex. *Proc. Natl Acad. Sci. USA* **108**, 16086–16091 (2011).
- Alexander, N. S. et al. Energetic analysis of the rhodopsin–G-protein complex links the α5 helix to GDP release. *Nature Struct. Mol. Biol.* **21**, 56–63 (2014).
- Neves, S. R., Ram, P. T. & Iyengar, R. G protein pathways. *Science* **296**, 1636–1639 (2002).
- Chothia, C. & Lesk, A. M. Helix movements and the reconstruction of the haem pocket during the evolution of the cytochrome c family. *J. Mol. Biol.* **182**, 151–158 (1985).
- Süel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Struct. Mol. Biol.* **10**, 59–69 (2003).
- del Sol, A., Fujihashi, H., Amorós, D. & Nussinov, R. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol.* **2** (2006).
- Kornev, A. P., Haste, N. M., Taylor, S. S. & Eyck, L. F. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl Acad. Sci. USA* **103**, 17783–17788 (2006).
- Zhang, X., Perica, T. & Teichmann, S. A. Evolution of protein structures and interactions from the perspective of residue contact networks. *Curr. Opin. Struct. Biol.* **23**, 954–963 (2013).
- Deupi, X. et al. Stabilized G protein binding site in the structure of constitutively active metarhodopsin-II. *Proc. Natl Acad. Sci. USA* **109**, 119–124 (2012).
- Standfuss, J. et al. The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471**, 656–660 (2011).
- Choe, H. W. et al. Crystal structure of metarhodopsin II. *Nature* **471**, 651–655 (2011).
- Scheerer, P. et al. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502 (2008).
- Slessareva, J. E. et al. Closely related G-protein-coupled receptors use multiple and distinct domains on G-protein α-subunits for selective coupling. *J. Biol. Chem.* **278**, 50530–50536 (2003).
- Oldham, W. M. & Hamm, H. E. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nature Rev. Mol. Cell Biol.* **9**, 60–71 (2008).
- Dratz, E. A. et al. NMR structure of a receptor-bound G-protein peptide. *Nature* **363**, 276–281 (1993).
- Tomba, P., Davey, N. E., Gibson, T. J. & Babu, M. M. A million peptide motifs for the molecular biologist. *Mol. Cell* **55**, 161–169 (2014).
- Sun, D. et al. Probing Gα₁₁ protein activation at single amino acid resolution. *Nature Struct. Mol. Biol.* (in the press).

26. Mnpotra, J. S. *et al.* Structural basis of G protein-coupled receptor-Gi protein interaction: formation of the cannabinoid CB2 receptor-Gi protein complex. *J. Biol. Chem.* **289**, 20259–20272 (2014).
27. Hanson, P. I. & Whiteheart, S. W. AAA+ proteins: have engine, will work. *Nature Rev. Mol. Cell Biol.* **6**, 519–529 (2005).
28. Nesbit, M. A. *et al.* Mutations affecting G-protein subunit $\alpha 11$ in hypercalcemia and hypocalcemia. *N. Engl. J. Med.* **368**, 2476–2486 (2013).
29. Iiri, T., Herzmark, P., Nakamoto, J. M., van Dop, C. & Bourne, H. R. Rapid GDP release from $G_{s\alpha}$ in patients with gain and loss of endocrine function. *Nature* **371**, 164–168 (1994).
30. Li, D. *et al.* Autosomal dominant hypoparathyroidism caused by germline mutation in GNA11: phenotypic and molecular characterization. *J. Clin. Endocrinol. Metab.* **99**, E1774–E1783 (2014).
31. de la Vega, M., Burrows, J. F. & Johnston, J. A. Ubiquitination: Added complexity in Ras and Rho family GTPase function. *Small GTPases* **2**, 192–201 (2011).
32. Oldham, W. M. & Hamm, H. E. Structural basis of function in heterotrimeric G proteins. *Q. Rev. Biophys.* **39**, 117–166 (2006).
33. Cui, Q. & Karplus, M. Allostery and cooperativity revisited. *Protein Sci.* **17**, 1295–1307 (2008).
34. Brown, C. J., Johnson, A. K., Dunker, A. K. & Daughdrill, G. W. Evolution and disorder. *Curr. Opin. Struct. Biol.* **21**, 441–446 (2011).
35. Tanaka, T. *et al.* α helix content of G protein or subunit is decreased upon activation by receptor mimetics. *J. Biol. Chem.* **273**, 3247–3252 (1998).
36. Natochin, M., Moussaif, M. & Artemyev, N. O. Probing the mechanism of rhodopsin-catalyzed transducin activation. *J. Neurochem.* **77**, 202–210 (2001).
37. Marin, E. P., Krishna, A. G. & Sakmar, T. P. Disruption of the $\alpha 5$ helix of transducin impairs rhodopsin-catalyzed nucleotide exchange. *Biochemistry* **41**, 6988–6994 (2002).
38. Marin, E. P., Krishna, A. G. & Sakmar, T. P. Rapid activation of transducin by mutations distant from the nucleotide-binding site: evidence for a mechanistic model of receptor-catalyzed nucleotide exchange by G proteins. *J. Biol. Chem.* **276**, 27400–27405 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank U. F. Lang, C. Chothia, R. Weatheritt, S. Balaji, H. Harbrecht, A. Morgunov, G. Murshudov and N. S. Latysheva for their comments on this work. This work was supported by the Medical Research Council (MC_U105185859; M.M.B., T.F., M.K., A.J.V. and C.N.J.R.; MC_U105197215; C.G.T.), the Swiss National Science Foundation grants 141898, 133810, 31-135754 (D.B.V.), the MRC Centenary Award (A.J.V.), the AFR scholarship from the Luxembourg National Research Fund (C.N.J.R.) and the Boehringer Ingelheim Fond (T.F.). M.M.B. is also a Lister Institute Research Prize Fellow.

Author Contributions T.F. collected data, wrote scripts and performed all the analysis. C.N.J.R. helped with data analysis and writing the scripts for structure analysis with T.F. A.J.V. helped T.F. with data analysis. D.S. and D.B.V. contributed experimental results on alanine scanning. M.K. prepared the CGN webserver. C.G.T. helped with aspects of data interpretation. T.F. and M.M.B. designed the project, analysed the results and wrote the manuscript. All authors read and provided their comments on the draft. M.M.B. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.F. (tflock@mrclimb.cam.ac.uk) or M.M.B. (madanm@mrclimb.cam.ac.uk).

METHODS

No statistical methods were used to predetermine sample size.

Generation of sequence and structure data sets. *Identification of relevant G α protein structures.* All structures related to the G α protein family (Pfam family: PF00503) were collected from Pfam (release 27.0) and Ensembl³⁹ using the R BioMart interface⁴⁰. In addition, the identified 973 G α homologue sequences (see below) were scanned against the entire Protein Data Bank (PDB) database using the BLAST algorithm⁴¹ to ensure all G α -containing structures were identified. 91 PDB entries (146 G α chains) were identified, of which two were obsolete (2pz3 retracted, 2ebc superseded by 3umr). Crystallographic data and coordinate files were retrieved from the RCSB PDB API (Tuesday 4 February 2014 at 16.00 PST). G α structures from the parasite *Entamoeba histolytica* (4fid) and *Arabidopsis thaliana* (2xtz), as well as non-full-length G α (1a9g and 1lvz are solution NMR studies of the C-terminal helix of G α , 3rbq contains the 11 amino acid N-terminal part of transducin bound to UNC119), were excluded from the analyses. Four structures of the last 10 C-terminal amino acid residues of G α_i bound to rhodopsin (2x72, 3dqb and 3pqr) or meta-rhodopsin (4a4m) were used for the GPCR–G α interface analysis. Five PDB entries had no publication associated and were manually traced back to their original articles: 3umr was published in Johnston *et al.*⁴² and 4g5o, 4g5q, 4g5r, 4g5s were discussed in a study by Jia *et al.*⁴³. The final set of structures in our analyses span orthologues from human, mouse, rat and cow and encompass twelve different G α genes from eight different G α subfamilies (GNAI1, GNAI3, GNAO, GNAS2, GNAT1, GNAQ, GNA12, GNA13), thereby representing all G α families (G α_s , G α_i , G α_q and G α_{12}). A full list of all retrieved PDBs is provided in Supplementary Table 1.

Identification of canonical human G α protein sequences and paralogue alignment. All relevant human G α protein isoforms and variants were obtained from Ensembl³⁹ using R (full list in Supplementary Table 1). The ‘canonical’ protein sequences for each of the 16 human G α genes, as defined by Uniprot⁴⁴, were used as representative sequences for each human G α gene throughout this work. The sequences were aligned using MUSCLE⁴⁵ and were manually refined using the consensus secondary structure as a guide (see below). Phylogenetic relationships of G α were obtained from Treefam⁴⁶ (family TF300673). The cladogram of the 16 canonical human G α protein alignment was built with the Phylogeny.fr web service⁴⁷ using the PhyML v3.0 algorithm⁴⁸ with the SH-like Approximate Likelihood-Ratio Test using the Jones–Taylor–Thornton substitution matrix and TreeDyn⁴⁹ for visualization.

Orthologue alignments of one-to-one G α orthologues of 16 human G α genes. Phylogenetic relationships of G α sequences were collected from TreeFam⁴⁶, the Orthologous Matrix (OMA) database⁵⁰ and EnsemblCompara GeneTrees (Compara)⁵¹ using R scripts. Compara had the highest fraction of complete G α sequences for each human G α gene, except for G α_s , for which OMA had a better sequence coverage. In total, 973 genes from 66 organisms were used, of which 773 were one-to-one orthologues. To build an accurate, low-gap alignment of such a number of sequences, 16 independent orthologous alignments for each human G α gene were first created by aligning one-to-one orthologue groups using the PCMA algorithm⁵² followed by manual refinement. Subsequently, each orthologue alignment was cross-referenced to the CGN (see below) by referencing its respective human sequence to the human paralogue alignment. Conservation scores of each CGN position were calculated using both sequence identity and sequence similarity, based on the BLOSUM62 substitution matrix (Supplementary Note) using all complete sequences of the cross-referenced alignments (561 sequences). Sequence conservation was mapped onto PDB structures (Supplementary Data) and visualized by generating PDB files with B-factors substituted by conservation scores.

Phylogenetic distances. The evolutionary distance of the retrieved sequences relative to human was evaluated with TimeTree⁵³. G α one-to-one orthologues extend back to chordate (sea squirts; *Ciona savignyi* and *Ciona intestinalis* for G α_{15}), separated around 722.5 million years from *Homo sapiens*, and the most ancestral one-to-many orthologue extends back to Opisthokonta (yeast; *Saccharomyces cerevisiae*), separated by 1,215 million years from human. In this work, we only investigated G proteins from organisms that have a GPCR–G-protein system. Since plants do not encode GPCRs and the heterotrimeric G proteins are known to be auto-activated, we did not consider the plant G proteins in our analysis.

Development of a common G α numbering (CGN) system. *Common G α numbering system.* Comparative analyses of different protein structures and sequences to infer general principles of a protein family require a way of relating structural, genomic, or experimental data from different studies to each topologically equivalent position on homologous proteins. For GPCRs, the Ballesteros–Weinstein numbering scheme⁵⁴ enables the referencing of positions in the transmembrane helices of different GPCRs, not considering loop regions. We sought to develop a common G protein numbering (CGN) system that includes loop regions and

describes G α residues in three levels of detail (DSP), similar to a postal address. D refers to the structural domain and is optional (catalytic GTPase domain: G; helical domain: H); S stands for one of the 37 consensus secondary structure elements (including loops) of the conserved G α topology; and P relates to the corresponding residue position within the consensus secondary structure element mapped to an alignment of all ‘canonical’ human G α sequences (Extended Data Fig. 1). For a detailed guide of how to use the CGN and map any G α protein, please refer to the CGN webserver (<http://www.mrc-lmb.cam.ac.uk/CGN>) and Supplementary Note.

Mapping structures to Uniprot sequences. Since several G α protein structures represent chimaeric G proteins, have peptide tags, or contain point mutations, each residue/position in a PDB structure was mapped to its Uniprot sequence(s) using the Structure Integration with Function, Taxonomy and Sequence (SIFTS)⁵⁵ webserver followed by a manual validation for missing positions. This allowed assigning residue positions of each G α structure to their equivalent positions in the human paralogue alignment and the orthologue alignments.

Determination of domain D and consensus secondary structure S. Secondary structure assignments were made for each G α structure using the STRIDE algorithm⁵⁶. The consensus secondary structure elements (SSEs) were determined by considering the most prominent secondary structure type at each topologically equivalent G α position when comparing the secondary structure assignment of all 80 G α structures (mean and standard deviation of secondary structure type at each CGN position were calculated). Topologically equivalent positions had a high agreement in their SSE assignment and showed well-defined flanking regions (Supplementary Note). In addition, the assigned consensus SSEs were manually confirmed through a 3D-structure alignment using MUSTANG⁵⁷, from which the domains (G-domain and H-domain) were defined. The G α SSE nomenclature follows a standardized expansion of the previously defined nomenclature⁵⁸: uppercase letters H and S represent helices or sheets, respectively. SSEs of the G domain follow a numerical identifier (H1, H2, ..., H5 and S1, S2, ..., S6 with the exception of HG and HN); SSEs of the H domain have an alphabetical identifier (HA, HB, ..., HF), starting from the N- to the C terminus of G α . The N-terminal region that forms a membrane-anchored helix is defined as HN. Systematic identifiers for historical names of some loop regions (switch regions, P-loop, etc.) were derived by concatenating the flanking SSE names using lowercase letters; for instance s6h5 refers to the loop between S6 and H5. A reference table including the historical loop names is provided in Extended Data Fig. 1 and Supplementary Table 2.

Determination of position P. P describes the amino acid position within an SSE, as determined by mapping the consensus secondary structure to the human paralogue alignment (Extended Data Fig. 1 and Supplementary Table 2). Insertions in orthologues are annotated P-i, where i stands for the number of inserted residues after position P, for instance Arg334^{H4.27-2} for the second amino acid of an insertion after position 27 of helix H4, found in pufferfish (*Tetraodon nigroviridis*) G α_s (Supplementary Note).

Consensus non-covalent contact networks between conserved residues. *Non-covalent residue contact networks.* Non-covalent contacts between residues of a protein define its topology, conformation and stability. For each of the 80 G α protein structures, a local version of the RINerator 0.5 package from 2014⁵⁹ was used to calculate H-bonds and van der Waals interactions between residues. Matrices of the all-against-all atomic distances of all residue contacts within each structure were computed using R and the bio3d package⁶⁰. Non-canonical interaction such as π – π stacking were identified with NCI⁶¹. All other calculations, analysis and processing were performed using custom written scripts in R.

Assignment of G α structures to signalling states. Structural differences between G α seem to arise from a convolution of the conformational state, binding partner, and G α protein type and species (Supplementary Note). To identify the non-covalent contacts of a structure that are crucial for each signalling state, and independent of the G α protein type and species, all G α structures were assigned to one of the four different G α signalling states depending on (1) the bound ligand, and (2) the interaction partner (Supplementary Table 1). The four states are (1) heterotrimeric GDP-bound state (inactive state), (2) nucleotide-free heterotrimeric receptor bound complex (GEF-bound state), (3) GTP γ S and potentially ‘effector’-bound state (active state), and (4) RGS-bound GDP+ALF hydrolysis transition state (GAP-bound state). Eleven structures are in the inactive state, one full-length structure (and four structures of the C-terminal G α peptide in complex with a GPCR) in the GPCR-bound state, 25 have GTP γ S bound or/and are co-crystallized with their downstream effectors, and 40 structures have G α in the GTP-hydrolysis transition state with GDP and aluminium fluoride (ALF) bound (GDP+ALF) and/or are co-crystallized with their RGS or a GTP-hydrolysis promoting peptide mimicking the RGS binding interface (for example, Go-Loco motif). Two structures (1cip, 1svs) had non-standard G α ligands bound, and 2zjz⁶² did not have a detailed description of its biochemical

relevance, and thus were not assigned to any signalling state. Eleven structures were identified as chimaeras and 21 included mutations (Supplementary Note). The publications describing the protein structure of each PDB entry were checked to confirm the relevance of the assigned signalling state.

Consensus contacts between conserved residues. To compare residue contact networks (RCNs) from different structures, topologically equivalent positions were cross-referenced with the CGN system. All RCN analyses, consensus RCN calculation, and conservation analysis were conducted using customized R scripts: matrices representing the absence or presence of non-covalent contacts between each possible pair of CGN residues in each PDB structure were computed (Supplementary Fig. 1). The consensus contacts of each signalling state were computed as the probability to find a contact in all structures of the state. Structure models can differ in their number of equivalent residues due to missing electron density, not fully fitted models, or truncations for crystallographic purposes. Thus, each consensus contact probability was normalized by the number of structures of the state that have the respective residue pair, in order to distinguish the absence of a contact from the absence of an equivalent position in a single PDB. To expand the structural analysis to other G α proteins for which only sequence data was available, sequence conservation was mapped to each CGN residue (see above).

Visualization of consensus contacts and identification of universal structural motifs. The consensus contacts between conserved residues in the different signalling states were visualized to investigate the contact re-organization in detail. For 2D visualization, the respective consensus RCNs were exported to Cytoscape⁶³ using the RCytoscape interface⁶⁴. For 3D visualization, R was used to create consensus RCNs in PyMol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.) by creating pseudo PDB structure coordinate files that show residues as spheres from their C-alpha atoms and lines/edges between them using CONECT entries. Information on sequence conservation was mapped via the B-factor field of the pseudo PDB structures. For simplification, only contacts present in more than 90% of all structures with a sequence identity >90% were shown as 'consensus contacts' between conserved residues—this threshold was chosen based on the bimodal distribution of contact occurrence (Supplementary Fig. 2). In addition, only long-range interactions (>i + 4) are shown for the consensus RCNs. It is important to note that these cut-offs were applied only for visualization purposes, while for the analysis, no cut-off was needed. All relevant consensus contacts were additionally visually inspected for each of the 80 PDB structures by creating automated PyMol sessions from R that superimpose all the 80 structures. To generate RCNs between SSEs, the sum of all contacts of the respective SSE as defined by the consensus SSE of the CGN were computed. Chimera⁶⁵ was used to manually re-evaluate atomic contacts, and PyMol was used to create publication-quality images.

Interface analysis. Buried surface area and inter-G α -GPCR residue contact networks. Inter-chain RCNs between G α and the receptor (G α_s and β_2 AR chains A and R in 3sn6, G α_t C-terminal peptide and rhodopsin from chains B and A in 2x72, 3dqb, 3pqr, 4a4m) were calculated as described above. The buried surface area (BSA) was obtained from the PDBe PISA (Proteins, Interfaces, Structures and Assemblies)⁶⁶ XML repository and normalized by the accessible surface area for each residue position. BSA and G α -GPCR RCNs were mapped to the CGN and the Ballesteros-Weinstein numbering, respectively. Sequence conservation from 561 complete G α homologue sequences and 249 human non-olfactory class A GPCRs was mapped onto the interface to determine the conserved 'hotspot' residues in the interface and visualized in PyMol (Supplementary Data). The BSA histogram, the visualization of the residue interaction network per secondary structure elements, and the correlation of BSA per residue versus conservation were produced in R and ggplot2.

Force-field-based energy estimations. The per-residue energy contributions to G α monomer and G α -GPCR complex stability were calculated using FoldX 3.0, which uses energy terms weighted by empirical data from protein engineering experiments to provide a quantitative estimation of each residue's contribution to protein stability and protein complex stability (<http://foldx.crg.es/>). For the interface analysis, the 3sn6 structure was energy minimized with the FoldX 'repair pdb' function and subsequently, the per-residue energy contributions for both the G α_s - β_2 AR complex and the monomers in isolation were calculated using the FoldX 'sequence detail', 'analyse complex', and 'stability' functions at 298K, pH 7.0, and 0.05M ionic strength. The per-residue energy contributions to complex stability were calculated as the difference between the energy contributions of each residue in the monomer and complex ($\Delta\Delta G_{\text{interface}}$) and visualized with R (Extended Data Fig. 2a). For energy contributions of each residue within G α monomers (Extended Data Fig. 4b), the average energy contribution and standard deviation for each G α position was computed after running the FoldX 'stability' and 'sequence detail' functions at 298K, pH 7.0, and 0.05M ionic strength for each of the 79 non-complex structures.

Disorder propensity calculations for all G α homologue sequences and structures. The disorder propensity of each of the 561 complete G α homologue sequences was calculated with IUPred⁶⁷ (prediction-type-setting: 'short disorder'). The missing structure positions were identified with bio3d package⁶⁰ (Extended Data Fig. 2b).

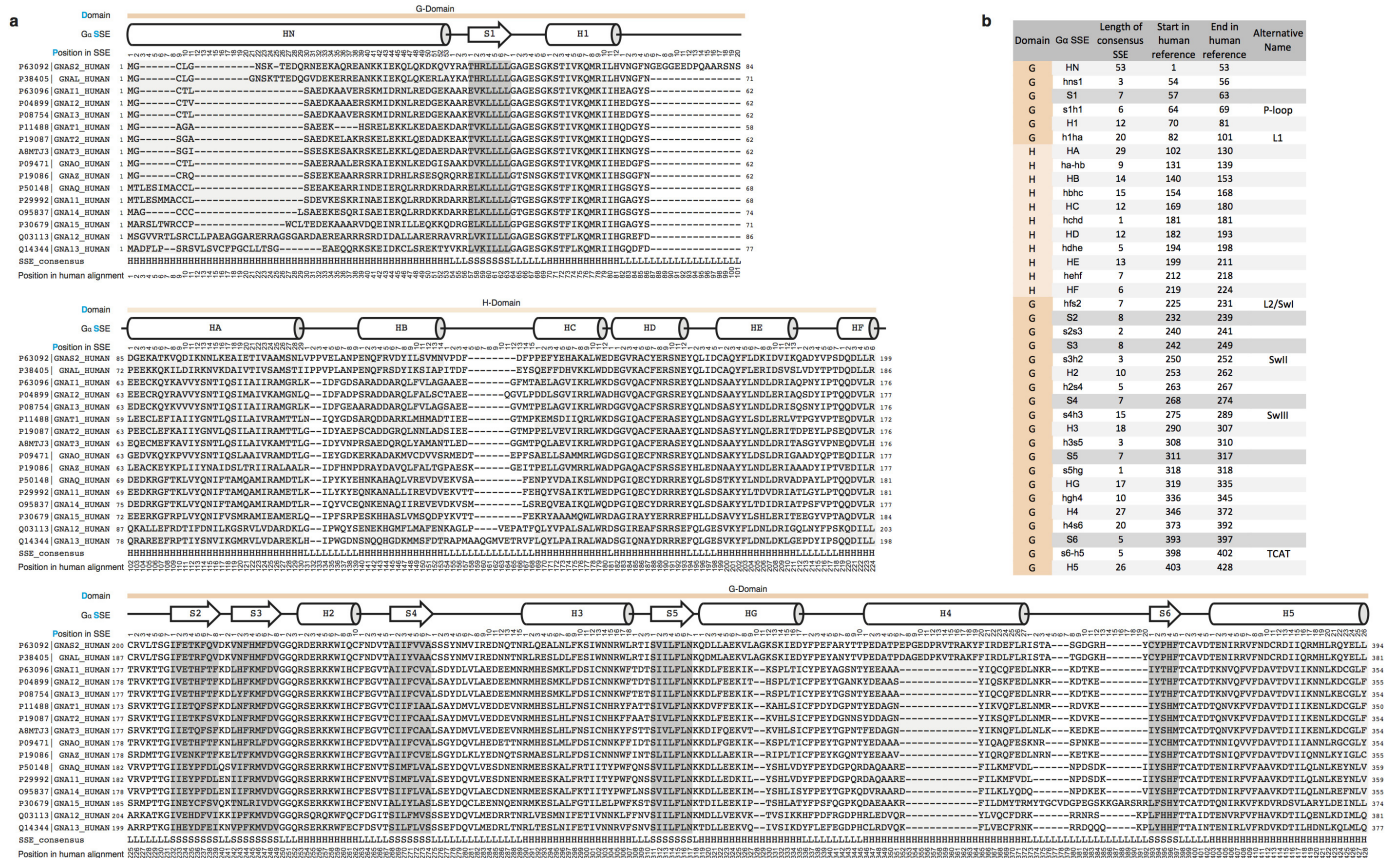
New and published mutational studies of different G α classes. Identification of mutations, mutant structures and chimaeras. Additional literature on G α mutations was retrieved with the text mining tool MutationMapper⁶⁸ and manually validated and filtered for correct hits/search results. Disease mutations were retrieved from the Database of Single Nucleotide Polymorphisms (dbSNP) and the Catalogue of Somatic Mutations in Cancer (COSMIC)⁶⁹ with biomaRt⁴⁰, and from the Human Gene Mutation Database (HGDM)⁷⁰. Mutations, chimaeras and peptide tags in the analysed structures were identified by comparing their Uniprot sequence to their PDB sequence using SIFTS⁵⁵ and mining PDBe annotations. All mutation data were mapped back to the CGN and visualized on their respective human G α structure.

Alanine scanning and stability of G α_t . The alanine scanning expression library of G α_{i1} was prepared as reported before⁷¹. The recombinant G α_{i1} alanine mutants were expressed in 24 well plates, purified by standard Ni-NTA affinity chromatography followed by buffer exchange using 96-well filter plates. The bovine rhodopsin and $\beta\gamma$ subunits were prepared from bovine retinas⁷². The melting temperature of each alanine mutant upon addition of GDP or GTP γ S was measured by differential scanning fluorimetry assay. The effect of each alanine G α_{i1} mutant on R*-G β complex formation and complex stability were measured by a high-throughput assay based on native gel electrophoresis. Detailed methods and protocols are provided in ref. 25.

G α versus Ras comparison. The Ras conformational cycle was featured in the RSCB PDB⁷³ April 2012 PDB-101 Molecule of the Month by David Goodsell (http://dx.doi.org/10.2210/rcsb_pdb/mom_2012_4), with high-resolution structures showing human HRas in its active GTP γ S-bound state (PDB 5p21⁷⁴) and the GDP-bound inactive state (PDB 4q21⁷⁵). 1bdk⁷⁶ was used as representative of the HRas GEF-bound state. These Ras representative structures were combined with an alignment of all human HRas paralogues identified from the OMA database⁵⁰. A structural alignment between the identified active and inactive Ras structures and the corresponding active and inactive G α (1got and 3ums) was used to accurately map Ras positions to the CGN (Supplementary Data) despite the low sequence identity (<6%) between Ras and G α . The number of atomic non-covalent contacts between helix H5 and helix H1 in the Ras and G α structures was manually compared in Chimera for the structures 1got and 4q21.

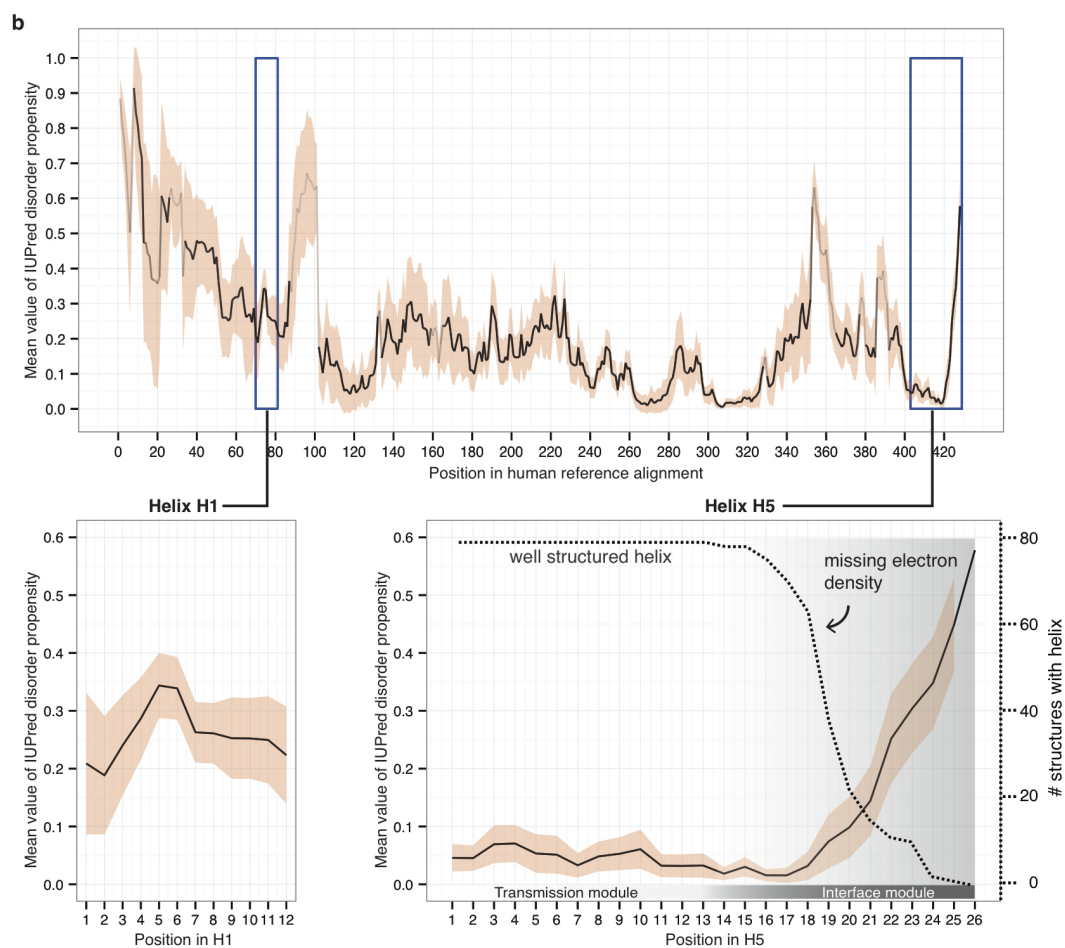
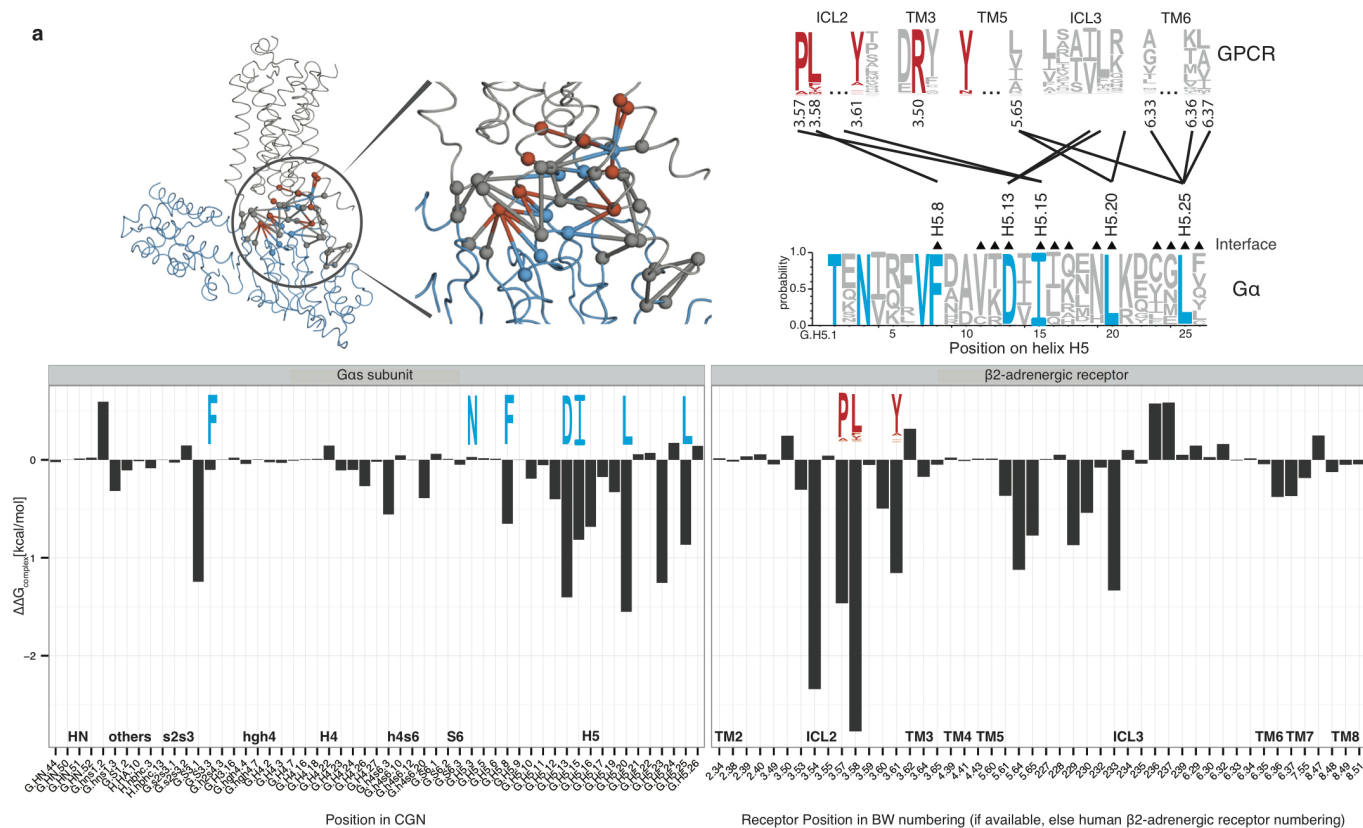
39. Flicek, P. et al. Ensembl 2014. *Nucleic Acids Res.* **42**, D749–D755 (2014).
40. Kasprzyk, A. BioMart: driving a paradigm change in biological data management. *Database* **2011**, bar049 (2011).
41. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
42. Johnston, C. A. et al. Structural determinants underlying the temperature-sensitive nature of a G α mutant in asymmetric cell division of *Caenorhabditis elegans*. *J. Biol. Chem.* **283**, 21550–21558 (2008).
43. Jia, M. et al. Crystal structures of the scaffolding protein LGN reveal the general mechanism by which GoLoco binding motifs inhibit the release of GDP from G α . *J. Biol. Chem.* **287**, 36766–36776 (2012).
44. Magrane, M. & Consortium, U. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* **2011**, bar009 (2011).
45. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
46. Ruan, J. et al. TreeFam: 2008 Update. *Nucleic Acids Res.* **36**, D735–D740 (2008).
47. Dereeper, A. et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).
48. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
49. Chevenet, F., Brun, C., Banuls, A. L., Jacq, B. & Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinform.* **7**, 439 (2006).
50. Altenhoff, A. M., Schneider, A., Gonnet, G. H. & Dessimoz, C. OMA 2011: orthology inference among 1000 complete genomes. *Nucleic Acids Res.* **39**, D289–D294 (2011).
51. Vilella, A. J. et al. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* **19**, 327–335 (2009).
52. Pei, J., Sadreyev, R. & Grishin, N. V. PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics* **19**, 427–428 (2003).
53. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2972 (2006).
54. Ballesteros, J. A. & Weinstein, H. *Receptor Molecular Biology* Vol. 25 (Elsevier, 1995).
55. Velankar, S. et al. SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Res.* **41**, D483–D489 (2013).

56. Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**, W500–W502 (2004).
57. Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J. & Lesk, A. M. MUSTANG: a multiple structural alignment algorithm. *Proteins* **64**, 559–574 (2006).
58. Noel, J. P., Hamm, H. E. & Sigler, P. B. The 2.2 Å crystal structure of transducin- α complexed with GTP γ S. *Nature* **366**, 654–663 (1993).
59. Doncheva, N. T., Klein, K., Domingues, F. S. & Albrecht, M. Analyzing and visualizing residue networks of protein structures. *Trends Biochem. Sci.* **36**, 179–182 (2011).
60. Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A. & Caves, L. S. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 (2006).
61. Babu, M. M. NCI: A server to identify non-canonical interactions in protein structures. *Nucleic Acids Res.* **31**, 3345–3348 (2003).
62. Morikawa, T. *et al.* Crystallization and preliminary X-ray crystallographic analysis of the receptor-uncoupled mutant of Gx1. *Acta Crystallogr. F* **63**, 139–141 (2007).
63. Lopes, C. T. *et al.* Cytoscape Web: an interactive web-based network browser. *Bioinformatics* **26**, 2347–2348 (2010).
64. Shannon, P. T., Grimes, M., Kutlu, B., Bot, J. J. & Galas, D. J. RCytoscape: tools for exploratory network analysis. *BMC Bioinform.* **14**, 217 (2013).
65. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
66. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
67. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434 (2005).
68. Vohra, S. & Biggin, P. C. Mutationmapper: a tool to aid the mapping of protein mutation data. *PLoS ONE* **8**, e71711 (2013).
69. Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).
70. Stenson, P. D. *et al.* The Human Gene Mutation Database: 2008 update. *Genome Med* **1**, 13 (2009).
71. Sun, D. *et al.* AAScan, PCRdesign and MutantChecker: a suite of programs for primer design and sequence analysis for high-throughput scanning mutagenesis. *PLoS ONE* **8**, e78878 (2013).
72. Maeda, S. *et al.* Crystallization scale preparation of a stable GPCR signaling complex between constitutively active rhodopsin and G-protein. *PLoS ONE* **9**, e98714 (2014).
73. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
74. Pai, E. F. *et al.* Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. *EMBO J.* **9**, 2351–2359 (1990).
75. Milburn, M. V. *et al.* Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science* **247**, 939–945 (1990).
76. Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D. & Kuriyan, J. The structural basis of the activation of Ras by Sos. *Nature* **394**, 337–343 (1998).



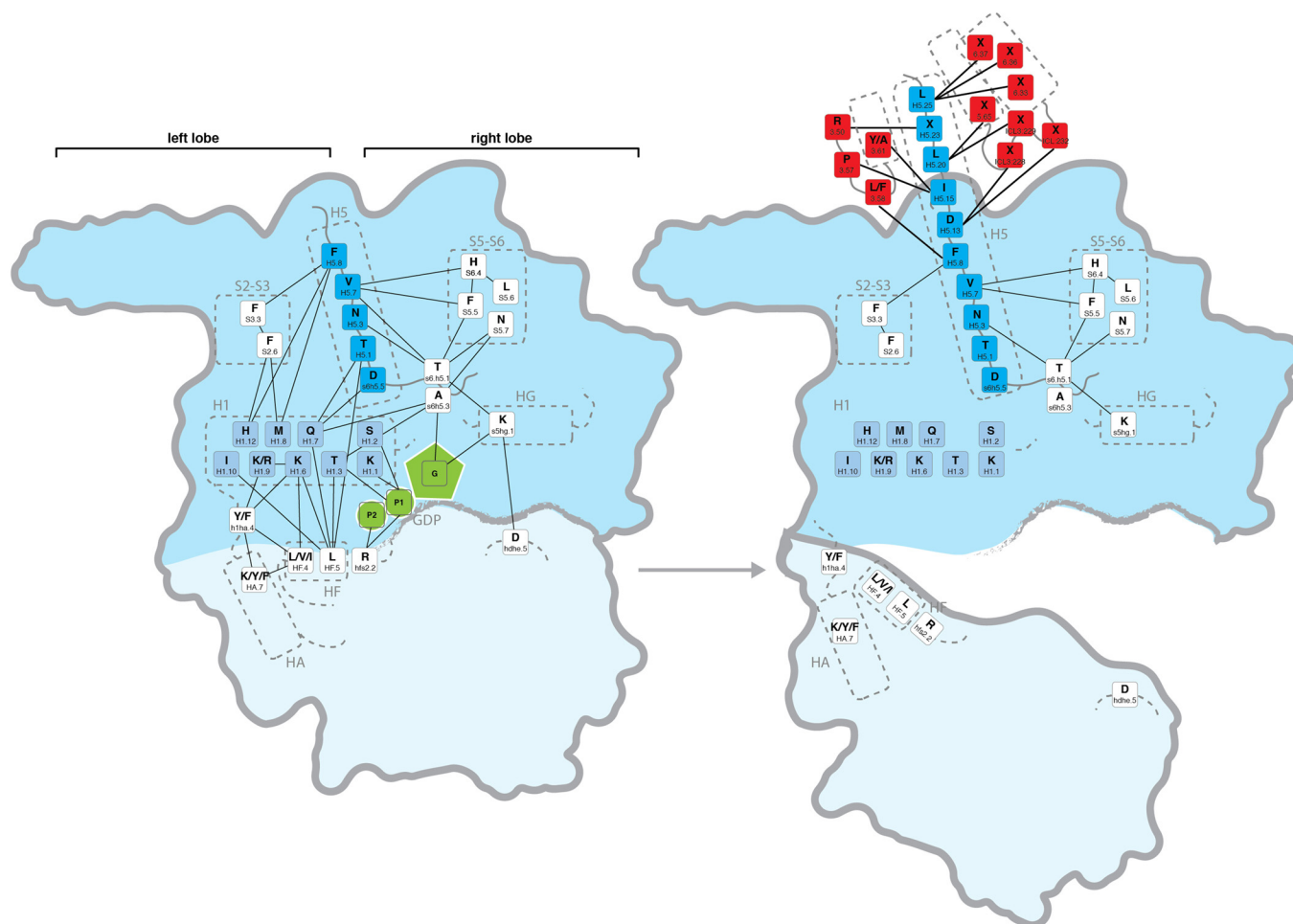
Extended Data Figure 1 | Human paralogue reference alignment for common $G\alpha$ numbering system. a, Reference alignment of all canonical human $G\alpha$ paralogues. The domain (D), consensus secondary structure (S) and

position in the SSE of the human reference alignment (P) are shown on top of the alignment. **b**, Reference table of the definitions of SSEs used in the CGN nomenclature.



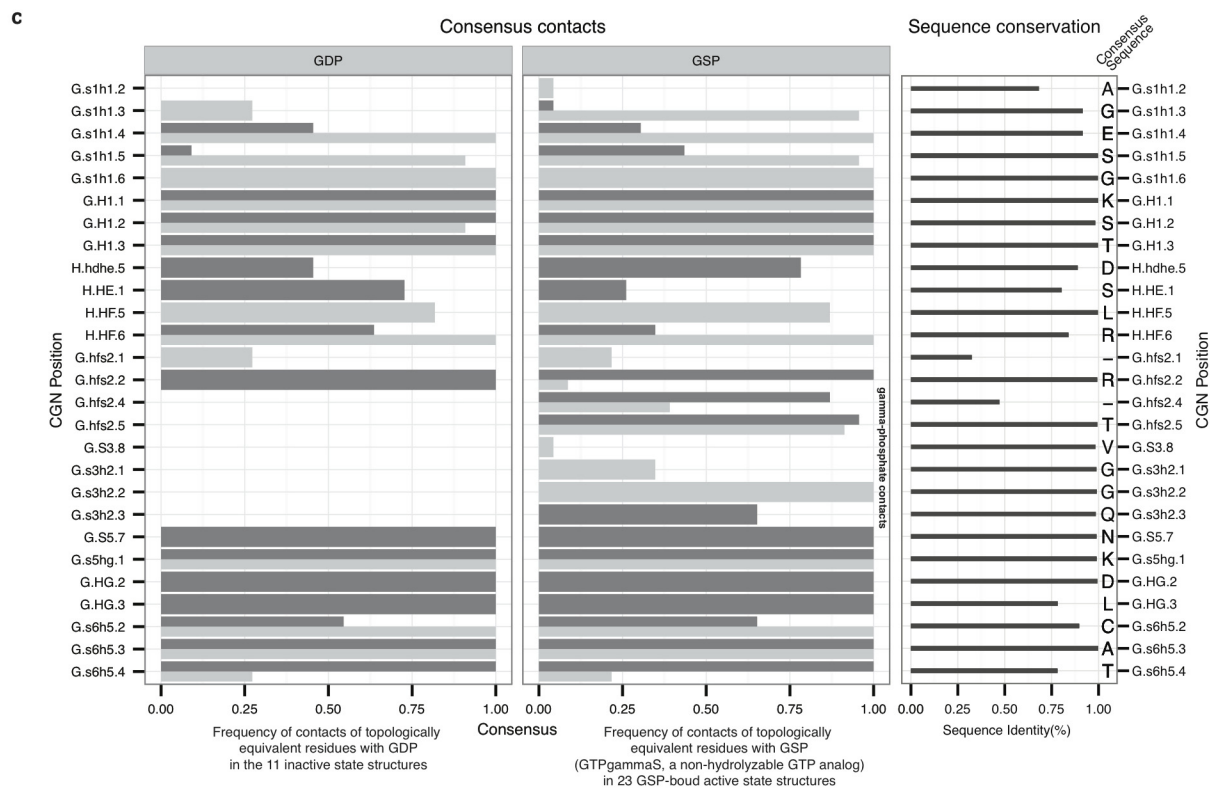
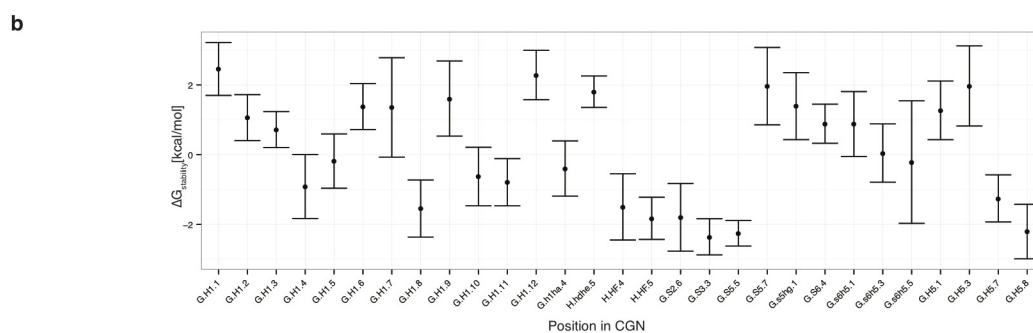
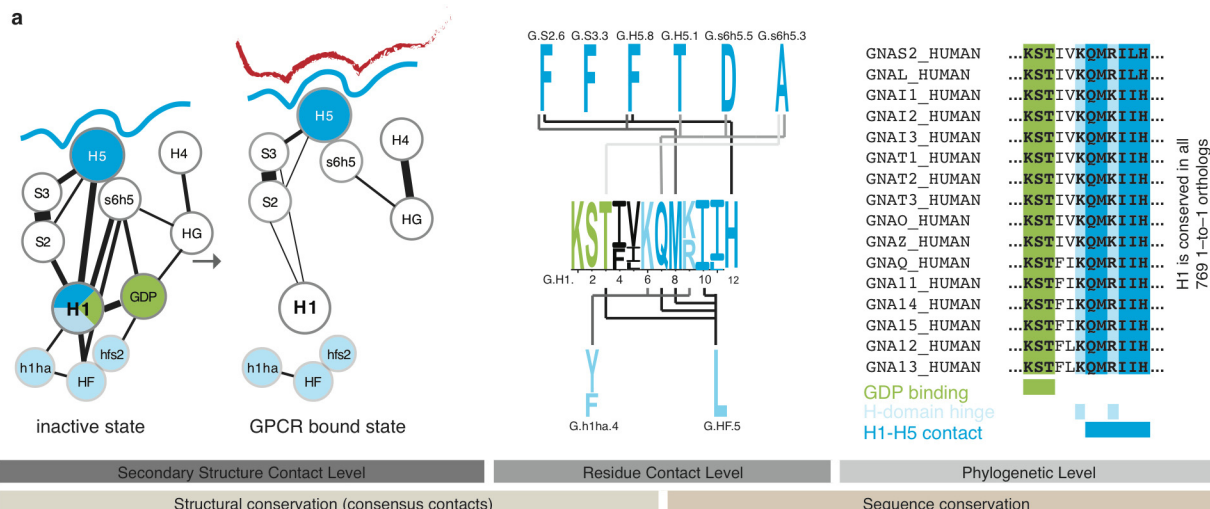
Extended Data Figure 2 | Energy estimation of the GPCR- $G\alpha$ residue contributions and $G\alpha$ disorder propensity. **a**, Energy contribution of single interface residues to the $G\alpha_s$ - β_2 AR complex calculated with FoldX ($T = 298\text{K}$, $\text{pH} = 7.0$, ionic strength = 0.05M). Conserved $G\alpha$ residues (blue sequence logo) that were identified to form receptor- $G\alpha$ inter-protein contacts with conserved GPCR residues (red sequence logo) are shown. The contact network between residues of the β_2 AR and $G\alpha_s$ is shown (red, conserved receptor residue; blue, conserved $G\alpha$ residue; grey, variable residues; spheres represent $C\alpha$ positions and links represent non-covalent contact). **b**, Consensus disorder

plot for all $G\alpha$ proteins. The mean value of the disorder propensity of all full-length $G\alpha$ sequences (561 sequences; homologous to all 16 human $G\alpha$ proteins) is shown as a black line; the standard deviation at each position is shown as light red ribbon. The colour tone of the line indicates the number of gaps at an aligned position (black, no gaps). The left inset shows the disorder propensity of H1. The right inset highlights that H5 is highly structured in its N terminus, and has increased disorder propensity towards the C terminus, which is in agreement with the missing electron density in the 79 structures.



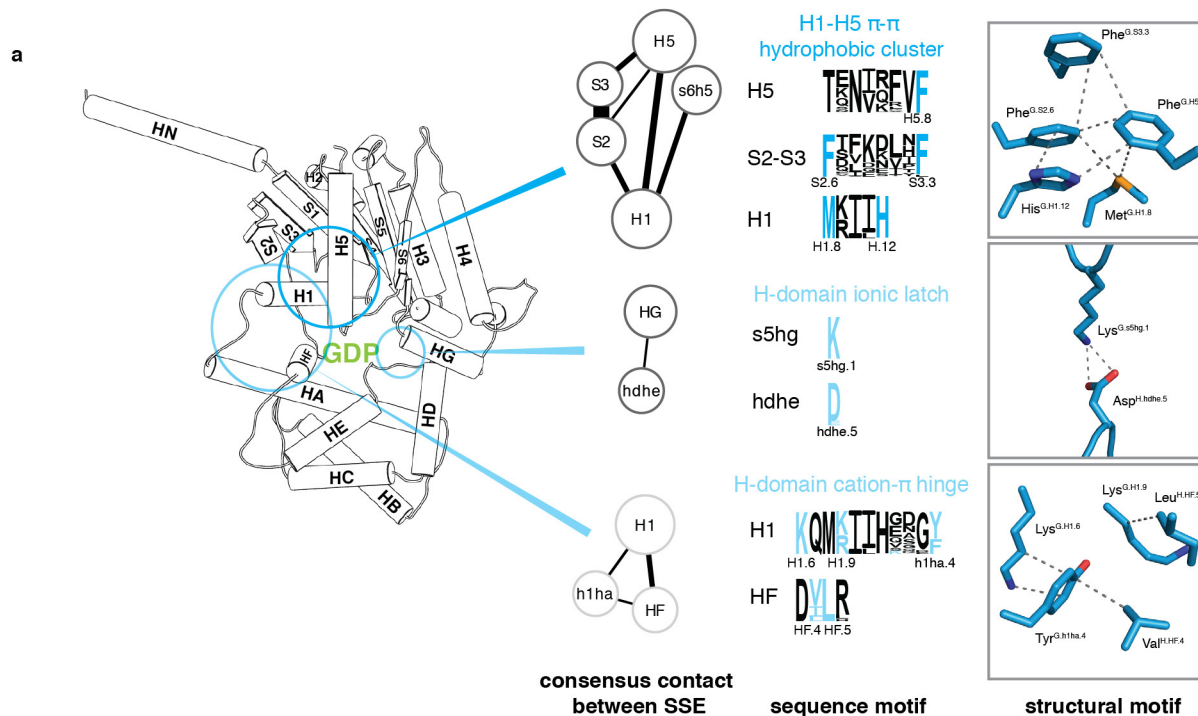
Extended Data Figure 3 | Rewiring of consensus contacts between conserved Gα residues upon receptor binding. CGN numbers and sequence logo for consensus contacts within Gα in the inactive state (left) and GPCR-bound state (right) are shown. Receptor residues are shown in red; H5 residues in dark blue; H1 residues in light blue; and GDP in green. The domains are

highlighted with a blue background (G-domain darker blue, H-domain light blue). This figure highlights the most important consensus residue contacts between conserved residues. Additional contacts in the right lobe are discussed in the Supplementary Note and in ref. 25. For a full list of residue contacts, please refer to Supplementary Data.



Extended Data Figure 4 | Details of helix H1 linking H5, GDP and the H-domain. **a**, This figure expands Fig. 4 from the main text to provide residue-level details of the role of helix H1. Residues forming contacts with H5 are shown in blue, with the H-domain in light blue and with GDP in green. Non-covalent consensus contacts between universally conserved residues at the SSE level (left) and per residue-level (centre). Lines denote non-covalent contacts between residues. The degree of conservation is shown as sequence logo. Residues are numbered according to the CGN. Helix H1 is almost 100% conserved across all 16 G α types and forms three structural motifs for interactions with H5, the H-domain and GDP (right). **b**, Average per residue energy contribution to G α protein stability as calculated from 79 structures

from all four G α subfamilies in the non-receptor-bound signalling states using FoldX ($T = 298\text{K}$, $\text{pH} = 7.0$, ionic strength = 0.05M). The average energy contribution is shown as dots, the standard deviation as bars. **c**, Per residue detail of G α -GDP and G α -GSP (non-hydrolysable GTP analogue) consensus contacts. The bar-plot shows the frequency of finding a contact mediated by topologically equivalent positions with GDP/GSP. Number of side-chain and main-chain contacts are shown as dark grey and light grey bars, respectively. The degree of conservation of contacting residues (calculated from the 561 complete G α sequences) is represented in the right panel and the consensus sequence for each position is shown.

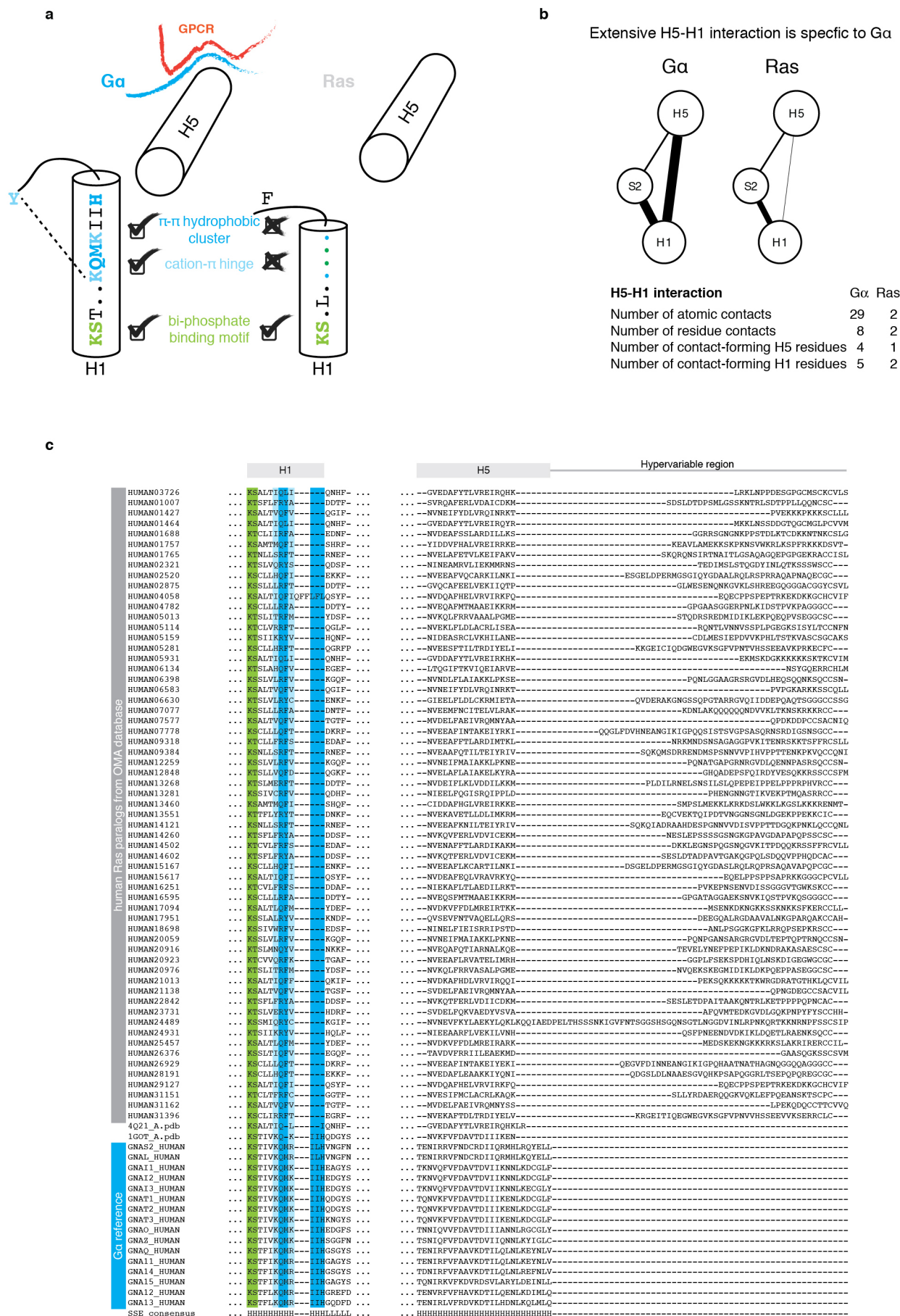


b

Position	Spontaneous activation	Disease	Binding vs. activation	G α type	Experiment	PMIDs
H5	✓	✓	✓	Gi	Changing H5 helicity induces GDP release	9452438
H5 modules	✓	✓	✓	Gt	Gly ₍₅₎ -spacer or Ala ₍₄₎ -spacer between H5 modules decouples G activation from GPCR binding	11279276, 12033931
Phe ^{G,H5.8}	✓	✓	✓	Gt, Gi, G11	Ala and Cys mutations accelerate GDP exchange, autosomal dominant hypocalcemia type 2	11356823, 24292645, 23802516
Thr ^{G,H5.1}	✓	✓	✓	Gt	Ala mutation causes constitutive GDP/GTP exchange	11356823
Val ^{G,H5.6}	✓	✓	✓	Gt	Ala mutation causes constitutive GDP/GTP exchange	11356823
Ala ^{G,H5.3}	✓	✓	✓	Gs	Ser mutation results in constitutive activation and causes testotoxicosis	8072545
Phe ^{G,S3.3}	✓	✓	✓	Gi	Cys mutation accelerates GDP exchange rate	24292645
Arg ^{G,H1.9}	✓	✓	✓	G11	Leu or Cys mutations cause autosomal dominant hypoparathyroidism	24823460, 23782177
Met ^{G,H1.8}	✓	✓	✓	Gi	Cys mutation accelerates GDP exchange rate	24292645
Ser ^{G,H1.2}	✓	✓	✓	Gi, Go	Cys mutation abolishes GTP-binding and phospholipase stimulation	7876152, 8408043
Gly ^{G,H5.2,7+G,H1ha.3}	✓	✓	✓	Gt/Gi chimera	Pro mutation causes constitutive GDP/GTP exchange	15271992

Extended Data Figure 5 | Conserved structural motifs of G α and known disease and engineered mutations. **a**, A universally conserved cluster of π - π and hydrophobic interactions between S2 (Phe^{G,S2.6}) and S3 (Phe^{G,S3.3}), H1 (Met^{G,H1.8} and His^{G,H1.12}) and H5 (Phe^{G,H5.8}) links H5 and H1 in the absence of the receptor. Upon receptor binding, residues within this motif (Phe^{G,H5.8} and Phe^{G,S3.3}) interact with the conserved Pro and Leu of ICL2 of the receptor as has been shown for G α_s (3sn6) and G α_i (ref. 26). Interrupting the contacts between H5 and H1 seems to be the trigger for transmitting the signal of GPCR binding to helix H1 (which interacts with GDP and the H-domain). The only conserved residue contact between the H-domain and the G-domain that is not in the hinge region is formed by a universally conserved salt bridge (H-domain ionic latch) between the very N-terminal end of HG of the G-domain (Lys^{G,s5hg.1}) and the loop connecting HD and HE in the H domain

(Asp^{H,hdhe.5}). The hinge region is formed by H1, the loop between H1 and HA, and HF. H1 interacts via (1) a cation- π interaction mediated by a universally conserved residue with the loop connecting H1 and HA (Lys^{G,H1.6} and Tyr^{G,h1ha.4}) and (2) a hydrophobic interaction with HF (Lys^{G,H1.9} and Leu^{H,HF.5}). **b**, Disease and engineered mutations that can be explained by the universal G α activation mechanism mapped on a G α protein. C α position of residues are shown as spheres; mutations at green positions cause spontaneous GDP release by interrupting consensus contacts between conserved residues, thereby 'mimicking' the effect of receptor binding to G α . Pink positions have also been reported to cause disease by constitutively activating G α . Insertion of an Ala₄ or Gly₅ after the yellow position separate the H5 transmission and interface module, thereby allowing GPCR binding without triggering GDP release.



Extended Data Figure 6 | Helix H5–H1 interaction in $G\alpha$ provides the allosteric GEF activation mechanism. **a**, Schematic representation of structural motifs on H1 that are shared or unique to $G\alpha$ and Ras. While the part of H1 with the phosphate-binding motif is conserved across both protein families, the C-terminal part is conserved only in $G\alpha$. H1 in $G\alpha$ has three additional residues that allow for extensive residue contacts between H1 and H5. In Ras, these interactions are missing and H5 and H1 are both 3 residues shorter. The consensus sequence and secondary structure of equivalent

residues of H1 in $G\alpha$ and Ras is also depicted. **b**, Comparison of the residue contact network between topologically equivalent residues in H5 and H1 in the corresponding inactive GDP-bound states of $G\alpha$ (1got) and Ras (4q21). The weight of the link between SSEs denotes the number of atomic contacts. **c**, Sequence alignments of H1 and H5 of human $G\alpha$ and Ras paralogues. The sequence alignment was obtained based on cross-referencing the alignments using the structures of $G\alpha$ and Ras.

Self-renewing diploid Axin2⁺ cells fuel homeostatic renewal of the liver

Bruce Wang^{1,2}, Ludan Zhao¹, Matt Fish¹, Catriona Y. Logan¹ & Roel Nusse¹

The source of new hepatocytes in the uninjured liver has remained an open question. By lineage tracing using the Wnt-responsive gene *Axin2* in mice, we identify a population of proliferating and self-renewing cells adjacent to the central vein in the liver lobule. These pericentral cells express the early liver progenitor marker *Tbx3*, are diploid, and thereby differ from mature hepatocytes, which are mostly polyploid. The descendants of pericentral cells differentiate into *Tbx3*-negative, polyploid hepatocytes, and can replace all hepatocytes along the liver lobule during homeostatic renewal. Adjacent central vein endothelial cells provide Wnt signals that maintain the pericentral cells, thereby constituting the niche. Thus, we identify a cell population in the liver that subserves homeostatic hepatocyte renewal, characterize its anatomical niche, and identify molecular signals that regulate its activity.

The cellular source of new hepatocytes in the adult liver and the molecular regulation of hepatocyte renewal are fundamental unanswered questions in liver biology. Recent studies in mice using genetic lineage tracing techniques have concluded that during homeostatic renewal, new hepatocytes arise by replication of pre-existing hepatocytes^{1,2}. This is in line with the generally accepted view that in the uninjured state, hepatocyte homeostasis does not involve a stem cell population³. However, hepatocytes are heterogeneous, with striking differences in age and function across the liver lobule⁴. In addition, mature hepatocytes are generally polyploid (4N to 32N), a genomic state that compromises replicative capacity^{5,6}, posing limitations on possible contributions of these cells to long-term liver homeostasis. It has been unknown whether a specific subpopulation of cells serves homeostatic renewal in the liver, as happens in many other tissues^{7–10}.

Wnt proteins are secreted short-range signals that maintain stem cells in many adult mammalian tissues, and are produced by the specialized microenvironment referred to as the stem cell niche¹¹. Wnt proteins signal primarily through the intracellular protein β -catenin to activate transcription. A universal transcriptional target of β -catenin-dependent Wnt signalling is *Axin2*, and its expression provides a reliable readout of cells responding to Wnt^{11,12}. Genetic lineage tracing of *Axin2*⁺ cells has identified stem cells in several adult mammalian tissues^{10,13}. We have used this lineage tracing approach to identify a unique population of Wnt-responsive cells that surround the central vein. These diploid cells self-renew over the lifespan of the animal and progressively give rise to mature polyploid hepatocytes that can populate the entire liver lobule. We also show that these pericentral cells are maintained by Wnt-producing central vein endothelial cells that constitute the niche.

Axin2⁺ pericentral cells generate expanding clones

In the adult liver, *Axin2* is expressed in cells located around the central vein^{14,15}, which we confirmed by *in situ* hybridization (Fig. 1m). In order to mark and follow the fates of these Wnt-responsive cells, we used the tamoxifen-inducible *Axin2*-CreERT2;*Rosa26*-mTmG^{fllox} mouse to pulse label *Axin2*⁺ cells. In these experiments, a subset of *Axin2*⁺ cells is labelled stochastically with membrane GFP after

tamoxifen administration. The GFP label is permanent, allowing for fate mapping of initially labelled cells and their descendants^{10,13}. A single low-dose of tamoxifen led to GFP labelling exclusively of pericentral hepatocytes (Fig. 1a). Control animals receiving corn oil did not show any GFP labelling (Extended Data Fig. 1). The GFP⁺ cells expressed glutamine synthetase (GS), another known Wnt target gene¹⁶ and a marker for pericentral hepatocytes (Fig. 1b). They were negative for carbamoyl-phosphate synthase 1 (CPS), which marks midlobular and periportal hepatocytes (Fig. 1c). Over time, the population of labelled cells expanded as large contiguous patches spreading directionally from the central vein towards the portal vein (Fig. 1d, g, j). One year after the marking, nearly all hepatocytes in some individual lobules were descendants of the initially labelled *Axin2*⁺ cells (Fig. 1j), including hepatocytes that abut the portal vein (Fig. 1j, inset).

Pericentral cells that remained labelled throughout the course of the lineage trace maintained their distinct gene expression profile, expressing *Axin2* (Extended Data Fig. 2) and GS (Fig. 1b, e, h, k) but not CPS (Fig. 1c, f, i, l). Conversely, the descendants of the labelled cells acquired different gene expression patterns as they moved away from the central vein. They lost *Axin2* and GS expression and gained CPS expression suggesting that, as they move away from the pericentral region, they no longer receive Wnt signals (see below) and subsequently differentiate. Finally, throughout the lineage traces, all labelled cells expressed the hepatocyte marker HNF4 α (Fig. 1n), but not markers of other liver cell types including biliary epithelial cells (data not shown), indicating that *Axin2*⁺ cells contribute only to the hepatocyte lineage.

While *Axin2*⁺ cells can generate all hepatocytes in a lobule over time, quantification of the labelling after one year showed that on average descendants of *Axin2*⁺ cells replaced 30% of the area of the entire liver (Fig. 1o and Extended Data Fig. 3), accounting for approximately 40% of the hepatocytes.

Axin2⁺ cells self-renew

A defining property of stem cells is the ability to self-renew. To test whether *Axin2*⁺ cells self-renew, we labelled a maximum number of *Axin2*⁺ cells by administering five consecutive daily doses of tamoxifen (Fig. 2a). Over time, the labelled cells expanded concentrically

¹Department of Developmental Biology, Howard Hughes Medical Institute, Stanford Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. ²Department of Medicine and Liver Center, University of California San Francisco, San Francisco, California 94143, USA.

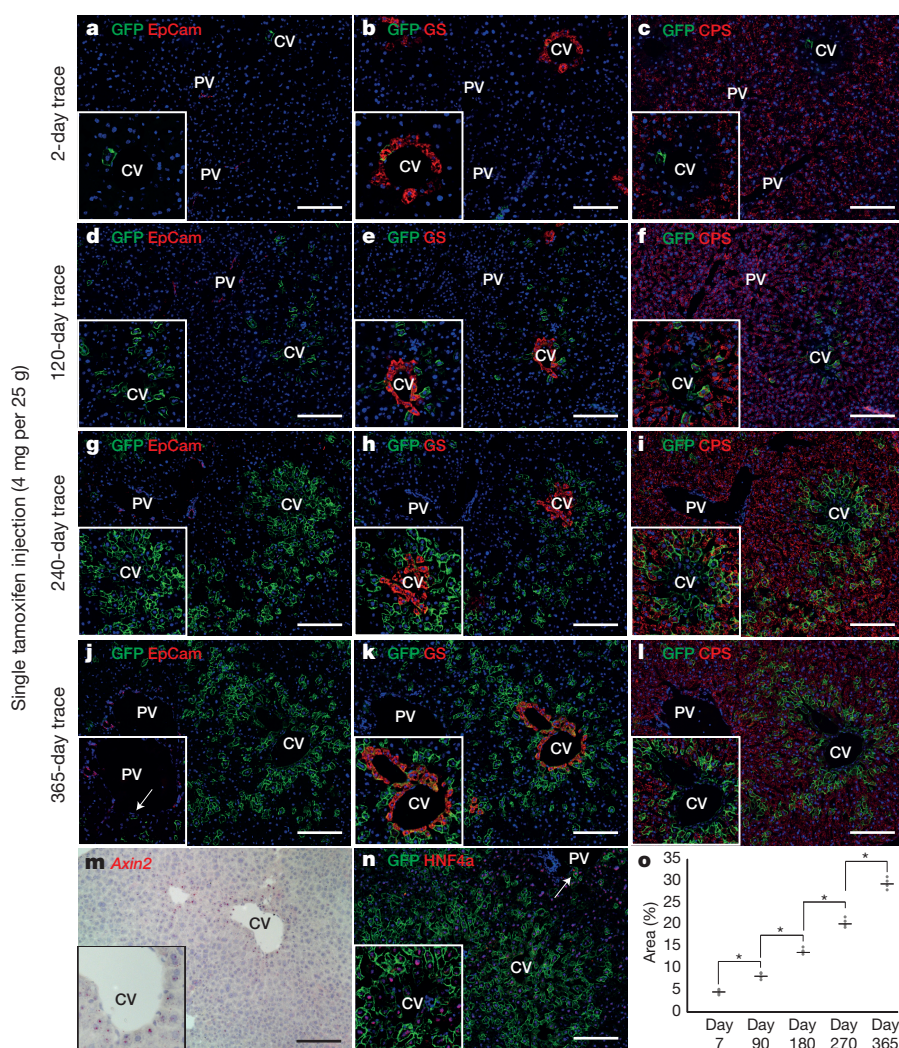


Figure 1 | Axin2⁺ pericentral cells generate expanding clones of hepatocytes from the central vein towards the portal vein over time. **a**, Few pericentral hepatocytes are labelled in Axin2-CreERT2;Rosa26-mTmG^{fllox} mice following a single dose of tamoxifen and traced for 2 days. EpCam labels bile ducts. **b**, **c**, Labelled pericentral cells express GS (**b**) but not CPS (**c**). **d**, **g**, **j**, The 120-day trace (**d**), 240-day trace (**g**) and 365-day trace (**j**) show expansion of labelled cells which can replace hepatocytes at the portal vein (**j** inset, arrow). **e**, **f**, **h**, **i**, **k**, **l**, Pericentral cells maintain GS expression (**e**, **h**, **k**), while labelled progeny acquire CPS expression (**f**, **i**, **l**). **m**, *In situ* hybridization for Axin2. **n**, All labelled cells express Hnf4a, including cells at the portal vein (arrow). **o**, Quantification of labelled hepatocytes over time. Data shows individual measurements and the mean. $n = 4$ animals for each time point. * $P < 0.05$, two-tailed unpaired *t*-tests. CV, central vein; PV, portal vein. Scale bars, 100 μ m.

from the central vein and, importantly, all pericentral cells remained labelled (Fig. 2b, c). This indicates that new pericentral cells arise exclusively from pre-existing labelled Axin2⁺ cells. Thus, while Axin2⁺ cells can give rise to all the hepatocytes along the lobule (Fig. 2c), they are not replaced by unlabelled Axin2⁻ cells. That is, pericentral Axin2⁺ cells are a self-renewing cell population.

To further characterize the Axin2⁺ cell population, we used RNA-seq to compare the gene expression profile of FACS-isolated Axin2⁺ and Axin2⁻ hepatocytes (Extended Data Fig. 4). As expected, most of the differentially expressed genes between the two populations were known markers of liver zonation (Extended Data Table 1)^{14,17}. However, we also identified *Tbx3*, a transcription factor important in maintaining pluripotency¹⁸, as a gene upregulated in the Axin2⁺ population. Notably, *Tbx3* marks early hepatoblasts that arise around

day 10 of embryogenesis, and is required for the earliest anlage of the liver¹⁹. By *in situ* hybridization, we confirmed that *Tbx3* expression is uniquely expressed in the single layer of pericentral cells (Fig. 2d).

In Axin2-CreERT2 mice, one copy of the endogenous *Axin2* gene is inactivated. Since *Axin2* is a negative feedback regulator of Wnt signalling¹², we considered the possibility that inactivation of one allele could confer a proliferative advantage after Wnt stimulation. To address this concern, we compared the DNA synthesis rate of Axin2⁺ hepatocytes in wild-type versus Axin2-CreERT2^{+/-} mice by 5-ethynyl-2'-deoxyuridine (EdU) incorporation. We used GS as the surrogate marker for Axin2⁺ cells in wild-type animals because a suitable antibody for murine liver Axin2 staining does not exist. We found no difference in the DNA synthesis rate in the two strains of mice (Extended Data Fig. 5). We also controlled for the possibility of

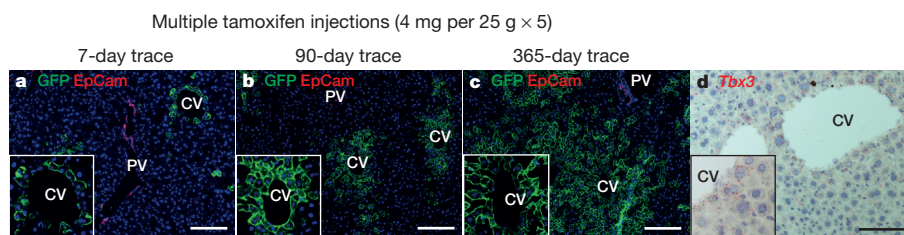


Figure 2 | Axin2⁺ cells self-renew. **a**, The majority of pericentral hepatocytes are labelled in Axin2-CreERT2;Rosa26-mTmG^{fllox} mice given five doses of tamoxifen and traced for 7 days. **b**, **c**, The 90-day trace (**b**) and 365-day

trace (**c**) show that all pericentral cells remain labelled (see insets). Note that non-labelled cells do not occupy the pericentral region over time. **d**, *In situ* hybridization of *Tbx3*. CV, central vein; PV, portal vein. Scale bars, 100 μ m.

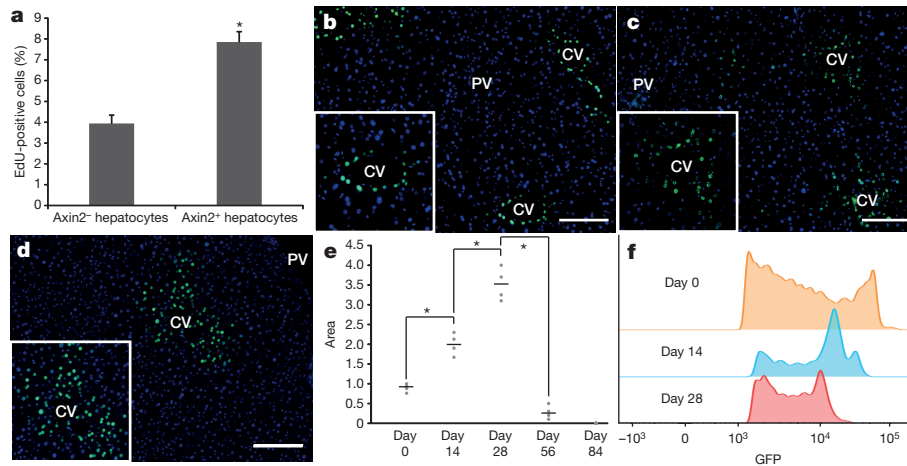


Figure 3 | Axin2⁺ hepatocytes proliferate faster than other hepatocytes. **a**, Quantification of EdU⁺ cells within Axin2⁻ and Axin2⁺ hepatocyte populations. Data represent mean \pm s.e.m. $n = 5$ animals. $*P < 0.05$, two-tailed unpaired t -test. **b**, All pericentral hepatocytes are labelled with nuclear GFP in Axin2-rtTA;TetO-H2B-GFP mice given doxycycline for 7 days. **c**, **d**, 14-

day chase (**c**) and 28-day chase (**d**) after doxycycline. **e**, Quantification of GFP-labelled nuclei. Data shows individual measurements and the mean. $n = 4$ animals per group. $*P < 0.05$, two-tailed unpaired t -tests. **f**, GFP intensity in day-0, day-14 and day-28 chase animals. Vertical axis shows number of events detected. CV, central vein; PV, portal vein. Scale bars, 100 μ m.

liver injury from tamoxifen administration, which could serve as a proliferative stimulus²⁰. We examined DNA synthesis by GS⁺ pericentral hepatocytes in mice treated with corn oil or tamoxifen, comparing wild-type and Axin2-CreERT2^{+/+} animals. We found no difference as measured by EdU incorporation (Extended Data Fig. 5). We conclude that neither *Axin2* gene dosage nor tamoxifen affects proliferation of Axin2⁺ cells in our mouse model.

Axin2⁺ cells proliferate faster than other hepatocytes

The fact that Axin2⁺ cells repopulate most of the liver lobule over time implies a rate of proliferation that is greater than that of Axin2⁻ hepatocytes. We quantified the DNA synthesis rates of the two cell populations as already described. Axin2-CreERT2;R26-mTmG^{fllox} mice were labelled with tamoxifen and then given seven daily doses of EdU. We found that Axin2⁺ cells undergo DNA replication twice as frequently as Axin2⁻ hepatocytes (Fig. 3a).

To verify this result, we also examined the replicative activity of Axin2⁺ cells by label dilution in which cells are tagged initially with a fixed amount of a stable product, which undergoes dilution with each cell division. We used the Axin2-rtTA;TetO-H2B-GFP transgenic mouse^{21,22}, which expresses a stable histone 2B-GFP fusion protein in Axin2⁺ cells when given doxycycline. Additionally, the reverse tetracycline-controlled transactivator in Axin2-rtTA mice is under the control of a mouse *Axin2* expression cassette²¹, thus leaving the endogenous *Axin2* gene locus unaffected. Once activated by doxycycline, the H2B-GFP protein remains stably expressed until the labelled cell undergoes cell division, when the H2B-GFP protein is divided between the daughter cells, resulting in diminished GFP signal intensity²².

Doxycycline was administered continuously for 7 days, at which time cells lining the central vein were labelled with nuclear GFP (Fig. 3b). Following a 14-day chase period after cessation of doxycycline, we observed that the number of GFP labelled cells had expanded, while the peak GFP signal intensity had decreased (Fig. 3c, f). The expansion of GFP-labelled cells around the central vein is concentric, consistent with the Axin2-CreERT2 lineage tracing results and is seen up to 28 days after cessation of doxycycline (Fig. 3d). By 56 days after cessation of doxycycline very few GFP-labelled cells are seen, and no labelled cells are observed after 84 days (Extended Data Fig. 6). We quantified the number of GFP⁺ cells at each time point and found that the cell cycling rate is approximately every 14 days (Fig. 3e). We further confirmed this by FACS analysis, which showed step-wise dilution of the peak GFP signal intensity every 14 days (Fig. 3f). Thus, pericentral Axin2⁺ cells actively proliferate

during adult homeostasis, at an estimated cell cycling rate of 14 days.

Axin2⁺ cells are mostly diploid

Mature hepatocytes are mostly polyploid²³, which is associated with decreased proliferative potential and increased senescence⁶. Stem cells are typically diploid, a property that may be necessary for unlimited duplication^{24,25}. We used FACS sorting to isolate Axin2⁺ cells and evaluated their ploidy status with Hoechst 33342 staining (Extended Data Fig. 7). As expected, the majority of unsorted hepatocytes are polyploid (Fig. 4a, c, left bar). In contrast, the majority of Axin2⁺ cells are diploid (Fig. 4b, c, middle bar). To confirm that Axin2⁺ diploid cells give rise to polyploid cells, we labelled Axin2⁺ cells with tamoxifen, traced them for one year and isolated GFP⁺ cells for ploidy analysis. We found that the ploidy distribution of the GFP⁺ cells at the end of the trace was identical to that of unsorted hepatocytes (Fig. 4c, right bar), suggesting that the descendants of Axin2⁺ cells mature normally into polyploid cells after leaving the pericentral zone.

Central vein endothelium acts as a Wnt-producing niche

The strict localization of Axin2⁺ cells to the central vein suggested a local source of Wnt. We screened the normal liver for all nineteen mammalian Wnts by *in situ* hybridization¹⁰. Two of these, *Wnt2* and *Wnt9b*, were expressed exclusively in endothelial cells around the central vein (Fig. 5a, b), co-localizing with the endothelial cell marker *Pecam1* (Fig. 5c, d). We also isolated liver endothelial cells by FACS²⁶ and confirmed that both *Wnt2* and *Wnt9b* are highly expressed in endothelial cells by quantitative reverse-transcription PCR (Fig. 5e and Extended Data Fig. 8). Thus endothelial cells at the central vein produce *Wnt2* and *Wnt9b* as short-range signals for pericentral Axin2⁺ cells and may constitute their niche.

Wnt signals are required for pericentral cell proliferation

The simultaneous and overlapping expression of two Wnt family members suggests functional redundancy of Wnt signalling at the central vein. To directly test whether endothelial-cell-derived Wnt proteins function to maintain the precursor state of pericentral cells, we conditionally deleted *Wntless* (*Wls*), a Wnt-specific transporter molecule required for proper Wnt protein secretion²⁷, specifically in endothelial cells. We crossed a VE-cadherin-CreERT2 mouse, which has a tamoxifen-inducible Cre-recombinase under the control of an endothelial cell specific promoter, with *Wls*^{fllox/fllox} mice. Adult VE-cadherin-CreERT2;*Wls*^{fllox/fllox} animals (*Wls*^{fl/fl}) were given multiple

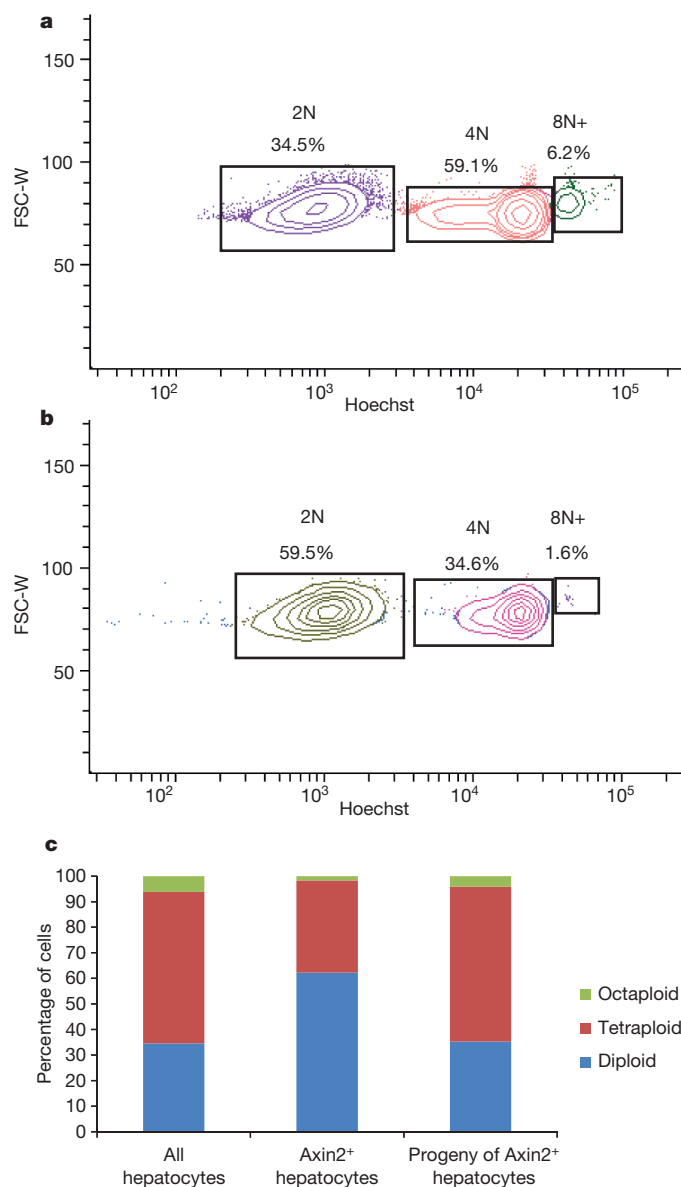


Figure 4 | Axin2⁺ hepatocytes are mostly diploid. **a**, FACS plot of hepatocytes stained with Hoechst 33342 and gated for diploid (2N), tetraploid (4N) and octaploid or greater (8N+) cells. **b**, FACS plot of Axin2⁺ hepatocytes stained with Hoechst 33342. FSC-W, forward scatter pulse width. **c**, Ploidy distribution within unsorted hepatocyte population (left), Axin2⁺ hepatocyte population (centre), or labelled hepatocytes after lineage tracing for one year (right). $n = 3$ animals per group.

doses of tamoxifen to induce conditional deletion of Wls in endothelial cells. These mice appeared healthy without obvious systemic defects and their livers appeared grossly and histologically normal (Extended Data Fig. 9). Compared to VE-cadherin-CreERT2;Wls^{fllox/+} (Wls^{fl/+}) control animals, Axin2 expression was decreased in pericentral hepatocytes of Wls^{fl/+} mice (Fig. 6a–c), consistent with loss of Wnt signalling around the central vein. Concurrently, there was loss of pericentral hepatocyte function, as shown by significantly decreased levels of GS expression (Fig. 6d–f). Importantly, pericentral cells, labelled by GS, exhibited significantly decreased proliferation rates in Wls^{fl/+} mice compared to Wls^{fl/+} controls (Fig. 6g). The decreased rate approached the proliferation rate of GS[−] hepatocytes, though it remained significantly higher. This is probably due to the mosaic nature of tamoxifen-induced Wls inactivation. We conclude that endothelial-cell-derived Wnt signals are necessary for maintaining the high proliferative state of pericentral cells.

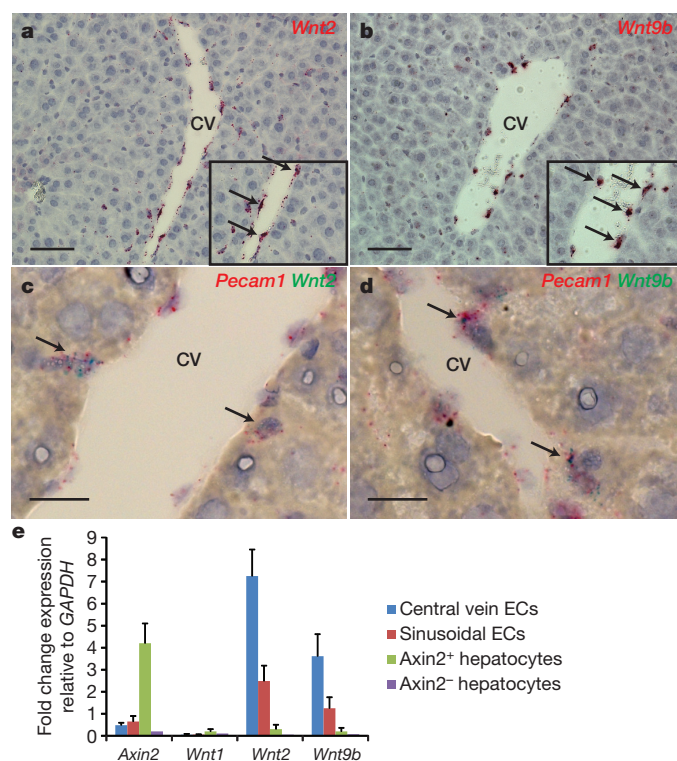


Figure 5 | Central vein endothelial cells produce Wnt proteins and act as a niche for pericentral cells. **a**, **b**, *In situ* hybridization of Wnt2 (**a**) and Wnt9b (**b**) showing mRNA expression at the central vein in endothelial cells (arrows, inset). **c**, **d**, Co-*in situ* hybridization of Pecam1 (red) and Wnt2 (green) (**c**) and Pecam1 and Wnt9b (**d**) showing co-expression in endothelial cells lining the central vein (arrows). **e**, Quantitative RT-PCR of Axin2, Wnt1, Wnt2 and Wnt9b of FACS-isolated liver cells. EC, endothelial cells. Data represent mean \pm s.e.m.; $n = 5$ animals. Scale bars, 100 μ m (**a**, **b**), 20 μ m (**c**, **d**).

Discussion

In this paper we present a new view of hepatocyte homeostasis in the uninjured liver (Fig. 6h). We have identified a Wnt-responsive cell population that resides within a confined niche around the central vein. These cells self-renew and contribute to hepatocyte maintenance by differentiating into and replacing other hepatocytes along the hepatic lobule in the normal liver. The existence of this pericentral cell population suggests that the fundamental mechanisms regulating liver renewal are similar to other organs in which homeostatic renewal involves small populations of stem cells that maintain the tissue. In the liver however, our model is novel because it was previously thought that all hepatocytes are equivalent in their renewal potential. In contrast, we show that hepatocytes are made up of more than one cell type and are not equivalent in replicative ability during homeostasis. Given the properties of the cell population under study, we postulate that the Wnt-responsive pericentral cells are hepatocyte stem cells.

Several features make pericentral cells unique compared to other hepatocytes. Although pericentral cells express markers common to other hepatocytes, they also specifically express Axin2, Tbx3 and GS while lacking CPS. Pericentral cells proliferate at a higher rate compared to other hepatocytes, an observation that is consistent with ref. 30. Furthermore, pericentral cells possess a diploid genome, in contrast to most other hepatocytes, which are polyploid. Finally, and most importantly, while pericentral Axin2⁺ cells can differentiate into all hepatocytes along the lobule, including those that line the portal vein, Axin2[−] hepatocytes do not replace pericentral cells during homeostasis. As pericentral cells can self-renew over the long term and differentiate into other hepatocytes, we suggest that they fit the functional definition of a stem cell.

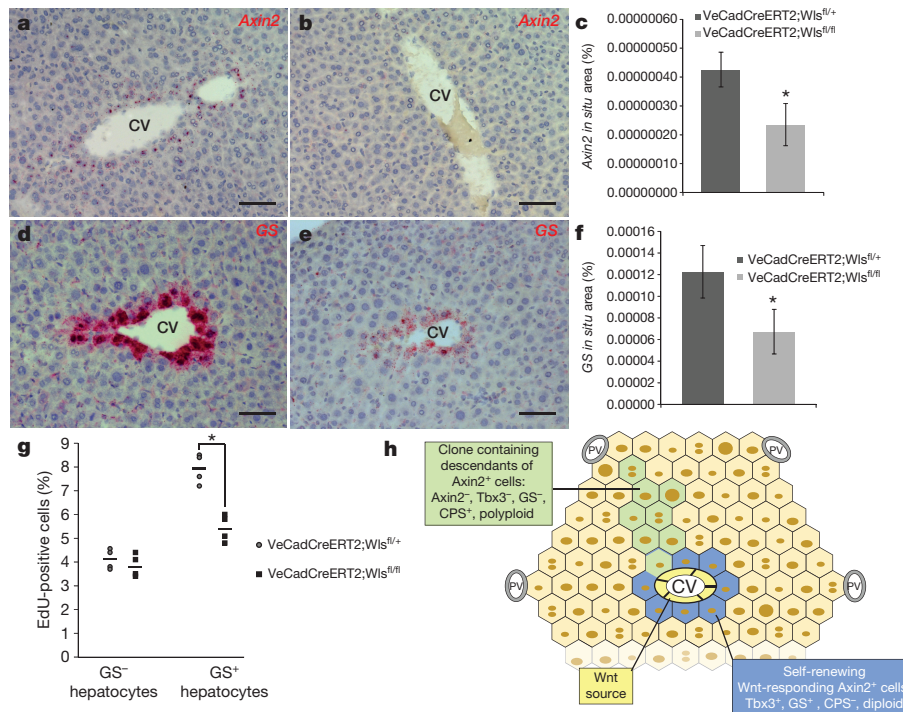


Figure 6 | Central-vein-derived Wnt proteins are required for pericentral cell proliferation. **a, b**, *Axin2* in situ hybridization in *Wls*^{fl/+} (**a**), and *Wls*^{fl/fl} (**b**) mice. **c**, Quantification of *Axin2* in situ signal. Data represent mean \pm s.e.m.; $n = 5$ animals per group. * $P < 0.05$, two-tailed unpaired *t*-test. **d, e**, *GS* in situ hybridization in *Wls*^{fl/+} (**d**) and *Wls*^{fl/fl} (**e**) mice. **f**, Quantification of *GS* in situ signal. Data represent mean \pm s.e.m.; $n = 5$ animals per group.

The diploid nature of pericentral cells is important and surprising, although nuclear size measurements in rat livers have suggested the presence of smaller nuclei near the central vein³¹. This sheds light on a long-standing question in liver biology. Mature polyploid hepatocytes display chromosomal abnormalities^{5,32} and display impaired replication^{5,6}. By maintaining a diploid genome, the pericentral cells would, like stem cells²⁵, retain unlimited replicative potential. It is interesting to note that during the cell cycle, levels of Wnt signalling peak at the G2/M phase³³. If Wnt proteins regulate expression of mitotic control genes such as the phosphatase *Cdc25*³⁴, they could direct cells to mitosis and continued diploidy rather than to non-mitotic DNA replication and polyploidy.

A defining feature of pericentral cells is their localization to a Wnt-rich anatomical niche. While Wnt-regulated genes such as β -catenin and *Apc* are known to function in liver development⁵⁵ and zonation¹⁴, the types and sources of Wnt have not been identified. We found that Wnt9b is specifically expressed in endothelial cells at the central vein, adjacent to the pericentral cells, while Wnt2 is expressed in both the sinusoidal and central vein endothelial cells. Notably, Wnt2 produced by sinusoidal endothelial cells is known to be important for hepatocyte regeneration after injury²⁶. Similarly, in other stem cell niches, lipid-modified Wnt signals act as short-range cues, maintaining stem cells in the immediate vicinity of the niche but not outside¹¹.

It has been suggested that there may be a periportal source of new hepatocytes under normal conditions³⁶. Our lineage tracing studies do not exclude the possibility that other sources of hepatocytes exist during homeostasis since after one year the descendants of pericentral cells replace on average only 40% of hepatocytes within the liver. However, a portal-based population would be regulated differently, since we find no expression of Wnt9b by the portal vein endothelium.

Liver is known to regenerate efficiently after injuries such as partial hepatectomy or chemical insult. It has been reported that during regeneration after chemical damage, a Wnt-responsive population

* $P < 0.05$, two-tailed unpaired *t*-test. **g**, Quantification of EdU⁺ cells within GS[−] or GS⁺ hepatocyte populations in *Wls*^{fl/+} and *Wls*^{fl/fl} animals. Data shows individual measurements and the mean. $n = 4$ animals per group. * $P < 0.05$, two-tailed unpaired *t*-test. **h**, Schematic of hepatocyte homeostatic renewal by pericentral cells. CV, central vein; PV, portal vein. Scale bars, 100 μ m.

of cells near the portal vein can be labelled by the *Lgr5* receptor gene¹⁵. These cells, unlike pericentral *Axin2*⁺ cells, do not express hepatocyte genes, but subsequently differentiate into bile duct epithelial cells and hepatocytes and thus could be similar to injury-induced oval cells³. Clearly, *Lgr5*⁺/oval cells are distinct from the cells we identify here, as pericentral cells maintain hepatocyte homeostasis in the uninjured liver while *Lgr5*⁺/oval cells have only been reported after injury.

The stem cell marker *Tbx3* is expressed widely in early liver hepatoblasts and is important for hepatoblast proliferation and initiation of hepatocyte differentiation^{19,37}. Our findings that pericentral cells also express *Tbx3* leads to the intriguing hypothesis that pericentral cells may represent the persistence of an embryonic hepatocyte progenitor population into a self-renewing cell population in the mature liver.

It is noteworthy that liver cancer is often characterized by loss-of-function mutations in negative regulators of the Wnt pathway, including *Axin* and *APC*³⁸. In a mouse model of liver cancer caused by *Met* overexpression, liver tumours were found to arise exclusively from cells located at the central vein^{39,40}, suggesting that pericentral *Axin2*⁺ cells, normally controlled by a paracrine Wnt signal, are precursors to liver cancer. This would explain why liver tumours contain mostly diploid cells⁴¹, an observation that was earlier rationalized by polyploid hepatocytes becoming diploid after oncogenic transformation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 November 2014; accepted 29 June 2015.

Published online 5 August 2015.

- Malato, Y. *et al.* Fate tracing of mature hepatocytes in mouse liver homeostasis and regeneration. *J. Clin. Invest.* **121**, 4850–4860 (2011).
- Yanger, K. *et al.* Adult hepatocytes are generated by self-duplication rather than stem cell differentiation. *Cell Stem Cell* **15**, 340–349 (2014).

3. Miyajima, A., Tanaka, M. & Itoh, T. Stem/progenitor cells in liver development, homeostasis, regeneration, and reprogramming. *Cell Stem Cell* **14**, 561–574 (2014).
4. Jungermann, K. & Kietzmann, T. Zonation of parenchymal and nonparenchymal metabolism in liver. *Annu. Rev. Nutr.* **16**, 179–203 (1996).
5. Ganem, N. J. & Pellman, D. Limiting the proliferation of polyploid cells. *Cell* **131**, 437–440 (2007).
6. Sigal, S. H. *et al.* Partial hepatectomy-induced polyploidy attenuates hepatocyte replication and activates cell aging events. *Am. J. Physiol.* **276**, G1260–G1272 (1999).
7. DasGupta, R. & Fuchs, E. Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. *Development* **126**, 4557–4568 (1999).
8. Zeng, Y. A. & Nusse, R. Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* **6**, 568–577 (2010).
9. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
10. Lim, X. *et al.* Interfollicular epidermal stem cells self-renew via autocrine Wnt signaling. *Science* **342**, 1226–1230 (2013).
11. Clevers, H., Loh, K. M. & Nusse, R. Stem cell signaling. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science* **346**, 1248012 (2014).
12. Lustig, B. *et al.* Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. *Mol. Cell. Biol.* **22**, 1184–1193 (2002).
13. van Amerongen, R., Bowman, A. N. & Nusse, R. Developmental stage and time dictate the fate of Wnt/ β -catenin-responsive stem cells in the mammary gland. *Cell Stem Cell* **11**, 387–400 (2012).
14. Benhamouche, S. *et al.* *Apc* tumor suppressor gene is the “zonation-keeper” of mouse liver. *Dev. Cell* **10**, 759–770 (2006).
15. Huch, M. *et al.* *In vitro* expansion of single *Lgr5*⁺ liver stem cells induced by Wnt-driven regeneration. *Nature* **494**, 247–250 (2013).
16. Cadoret, A. *et al.* New targets of β -catenin signaling in the liver are involved in the glutamine metabolism. *Oncogene* **21**, 8293–8301 (2002).
17. Braeuning, A. *et al.* Differential gene expression in periportal and perivenous mouse hepatocytes. *FEBS J.* **273**, 5051–5061 (2006).
18. Han, J. *et al.* *Tbx3* improves the germ-line competency of induced pluripotent stem cells. *Nature* **463**, 1096–1100 (2010).
19. Suzuki, A. *et al.* *Tbx3* controls the fate of hepatic progenitor cells in liver development by suppressing p19ARF expression. *Development* **135**, 1589–1595 (2008).
20. Moreira, P. I. *et al.* Estradiol affects liver mitochondrial function in ovariectomized and tamoxifen-treated ovariectomized female rats. *Toxicol. Appl. Pharmacol.* **221**, 102–110 (2007).
21. Yu, H. M. *et al.* Impaired neural development caused by inducible expression of Axin in transgenic mice. *Mech. Dev.* **124**, 146–156 (2007).
22. Tumber, T. Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004).
23. Guidotti, J. E. *et al.* Liver cell polyploidization: a pivotal role for binuclear hepatocytes. *J. Biol. Chem.* **278**, 19095–19101 (2003).
24. Comai, L. The advantages and disadvantages of being polyploid. *Nature Rev. Genet.* **6**, 836–846 (2005).
25. Ohlstein, B., & Spradling, A. The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. *Nature* **439**, 470–474 (2006).
26. Ding, B.-S. *et al.* Inductive angiocrine signals from sinusoidal endothelium are required for liver regeneration. *Nature* **468**, 310–315 (2010).
27. Bänziger, C. *et al.* Wntless, a conserved membrane protein dedicated to the secretion of Wnt proteins from signaling cells. *Cell* **125**, 509–522 (2006).
28. Monvoisin, A. *et al.* VE-cadherin-CreER^{T2} transgenic mouse: a model for inducible recombination in the endothelium. *Dev. Dyn.* **235**, 3413–3422 (2006).
29. Carpenter, A. C. *et al.* Generation of mice with a conditional null allele for Wntless. *Genesis* **48**, 554–558 (2010).
30. Magami, Y. *et al.* Cell proliferation and renewal of normal hepatocytes and bile duct cells in adult mouse liver. *Liver* **22**, 419–425 (2002).
31. Zajicek, G. & Schwartz-Arad, D. Streaming liver VII: DNA turnover in acinus zone-3. *Liver* **10**, 137–140 (1990).
32. Duncan, A. W. *et al.* The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature* **467**, 707–710 (2010).
33. Niehrs, C. & Acebron, S. P. Mitotic and mitogenic Wnt signalling. *EMBO J.* **31**, 2705–2713 (2012).
34. Vijayakumar, S. *et al.* High-frequency canonical Wnt activation in multiple sarcoma subtypes drives proliferation through a TCF/ β -catenin target gene, CDC25A. *Cancer Cell* **19**, 601–612 (2011).
35. Tan, X. *et al.* β -Catenin deletion in hepatoblasts disrupts hepatic morphogenesis and survival during mouse development. *Hepatology* **47**, 1667–1679 (2008).
36. Zajicek, G., Oren, R. & Weinreb, M. Jr. The streaming liver. *Liver* **5**, 293–300 (1985).
37. Lüdtke, T. H. *et al.* *Tbx3* promotes liver bud expansion during mouse development by suppression of cholangiocyte differentiation. *Hepatology* **49**, 969–978 (2009).
38. Laurent-Puig, P. & Zucman-Rossi, J. Genetics of hepatocellular tumors. *Oncogene* **25**, 3778–3786 (2006).
39. Wang, R. *et al.* Activation of the Met receptor by cell attachment induces and sustains hepatocellular carcinomas in transgenic mice. *J. Cell Biol.* **153**, 1023–1034 (2001).
40. Tward, A. D. *et al.* Distinct pathways of genomic progression to benign and malignant tumors of the liver. *Proc. Natl Acad. Sci. USA* **104**, 14771–14776 (2007).
41. Schwarze, P. E. *et al.* Emergence of a population of small, diploid hepatocytes during hepatocarcinogenesis. *Carcinogenesis* **5**, 1267–1275 (1984).

Supplementary Information is available in the online version of the paper.

Acknowledgements These studies were supported by the Howard Hughes Medical Institute and a grant from the Reed-Stinehart foundation. R.N. is an investigator with the Howard Hughes Medical Institute. B.W. was supported by F32DK091005. We thank D. M. Bissell and T. Desai for comments on the manuscript, V. Waehle for assistance in preparing RNA samples for RNA-seq, M. Britton for RNA-seq analysis, and P. Lovelace for assistance with FACS.

Author Contributions B.W. carried out the experiments. D.Z. performed qRT-PCR analysis. M.F. performed *in situ* hybridization. C.Y.L. performed RNA-seq analysis. B.W. and R.N. designed the study, analysed data and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE68806. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.N. (rnusse@stanford.edu) and B.W. (bruce.wang@ucsf.edu).

METHODS

Animals. B6 and FVB mice (Charles River Laboratories) were used for wild-type analysis. Axin2-CreERT2 mice¹³ and VE-cadherin-CreERT2 mice²⁸ have been previously described. Rosa26-mTmG^{lox} (*Gt(ROSA)26Sor^{tm4}(ACTB-tdTomato, EGFP)^{Luo}/J*)⁴², Wls^{lox} (129S-Wls^{tm1.1Lan}/J)²⁹, Axin2-rtTA (B6.Cg-Tg(Axin2-rtTA2S^{*}M2)7Cos/J)²¹, and TetO-H2B-GFP (Tg(tetO-HIST1H2BJ/GFP)47Efu/J)²² mice were obtained from The Jackson Laboratory. All alleles were heterozygous, except where stated.

For lineage tracing studies, Axin2-CreERT2;Rosa26-mTmG^{lox} mice 8–12 weeks of age received intraperitoneal injections of tamoxifen (Sigma, 4 mg per 25 g mouse weight) dissolved in 10% ethanol/corn oil (Sigma) either once or on five consecutive days. Representative figures of Axin2-CreERT2;Rosa26-mTmG^{lox} lineage tracing (Figs 1a–l and 2a–c) are from *n* = 5 animals per time point. Quantification of the area of labelled hepatocytes (Fig. 1o) was performed on Axin2-CreERT2/Rosa26-mTmG^{lox} mice given five daily doses of tamoxifen and lineage traced.

For label dilution studies Axin2-rtTA;TetO-H2B-GFP mice 8–12 weeks old received doxycycline hyclate (Sigma, 1 mg ml⁻¹) in drinking water for 7 days, then chased for 0, 14, 28, 56 or 84 days. Representative figures of Axin2rtTA;TetO-H2B-GFP label dilution (Fig. 3b–d) are from *n* = 4 animals per time point.

For the endothelial cell conditional knockout of Wntless, VE-cadherin-CreERT2;Wls^{lox} mice (Wls^{fl} and Wls^{fl/+}) aged 8–10 weeks received intraperitoneal injections of tamoxifen on five consecutive days and sacrificed 7 days after the last dose of tamoxifen. For proliferation studies (Fig. 6g), mice were given seven consecutive daily doses of EdU after the last dose of tamoxifen and sacrificed 2 h after the last EdU dose.

All animal experiments and methods were approved by the Institutional Animal Care and Use Committee at Stanford University. Mice used in this study were age- and gender-matched littermates including both sexes. All mice were housed in the animal facility of Stanford University on a 12-h light/dark cycle with *ad libitum* access to water and normal chow except when otherwise indicated. The animal experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Statistics. The statistical analysis used to measure significance is the two-tailed unpaired Student's *t*-test. The G*Power calculator (G*Power 3.1.9.2) was used for sample size calculations. For an α probability of 0.05, a β probability of 0.8, expected observed differences between the two comparison groups of 50% and assuming 15% standard error of the mean within each group, the sample size required to detect a statistically significant difference is four animals per group.

Liver histology and immunofluorescence. Mouse livers were fixed in 4% PFA overnight at 4 °C, cryoprotected in 30% sucrose for 24 h at 4 °C then embedded in OCT and snap frozen. All immunofluorescence staining was performed in the dark. Cryosections (10 μ m) were incubated in blocking buffer (5% normal donkey serum, 0.5% Triton-X in PBS) at room temperature and stained with primary and secondary antibodies, then mounted in Prolong Gold with DAPI mounting medium (Life Technologies). The following primary antibodies were used: GFP (chicken, 1:500, Abcam ab13970), GS (mouse, 1:500, Millipore MAB302), CPS (rabbit, 1:100, gift from W. Lamers), EpCAM (rabbit, 1:100, Developmental Studies Hybridoma Bank, clone g8.8), HNF4 α (rabbit, 1:100, Santa Cruz Biotechnology sc-8987).

Hepatocyte proliferation assay. Hepatocyte proliferation *in vivo* was measured by 5-ethynyl-2'-deoxyuridine (EdU) uptake. In brief, mice received a dose of intraperitoneal EdU (Life Technologies, 50 mg per kg mouse weight) daily for 7 days and harvested half a day after the final EdU dose. For EdU detection, cryosections were first stained with the appropriated primary and secondary antibodies, then incubated with the reagents in the Click-iT EdU Alexa Fluor 555 Imaging Kit (Life Technologies) prepared according to the manufacturer's instructions and mounted in Prolong Gold with DAPI mounting medium.

Liver cell isolation and flow cytometry. Hepatocytes and liver endothelial cells were isolated from mice by a two-step collagenase perfusion technique with modifications. In brief, after the inferior vena cava was cannulated and portal vein was cut, the liver was perfused at 10 ml min⁻¹ through the inferior vena cava with Liver Perfusion Medium (Invitrogen) at 37 °C for 10 min, followed by perfusion with collagenase type IV (Wellington) for an additional 10 min. The liver was dissociated and passed through a 70 μ m filter. Hepatocytes were separated from non-parenchymal cells (NPCs) by low-speed centrifugation (30g for 3 min \times 3), and further purified by Percoll gradient centrifugation as previously described⁴³. NPCs were pelleted from supernatant by centrifugation (300g for 5 min \times 3) and then stained with cell surface markers for endothelial cell isolation and flow cytometric analysis as previously described²⁶.

Cells were analysed on FACS ARIA II (BD). Data were processed with FACSDiva 8.0 software (BD) and FlowJo v10 (FlowJo). Doublets were excluded by FSC-W \times FSC-H and SSC-W \times SSC-H analysis. Single-stained channels were used for compensation and fluorophore minus one controls were used for gating.

The following antibodies were used: CD31-PE (eBioscience 12-0311-81), CD34-FITC (eBioscience 11-0341-85), rat VEGFR3 (gift from B.-S. Ding) with anti-rat Alexa Fluor 647 secondary (Jackson Immuno 712-605-153). Sinusoidal endothelial cells were identified as VEGFR3⁺CD31⁺CD34⁻ cells. Central vein endothelial cells were identified as VEGFR3⁻CD31⁺CD34⁺ cells.

Hepatocyte ploidy measurement. Hepatocyte ploidy staining was performed as previously described³². Wild-type FVB mice were used for baseline hepatocyte ploidy measurements. For ploidy measurement of Axin2⁺ hepatocytes, Axin2-CreERT2;Rosa26-mTmG^{lox} mice were given five daily doses of tamoxifen (4 mg per 25 g body weight) and cells were isolated 2 days or 1 year after the last dose of tamoxifen and stained with Hoechst 33342 (Invitrogen). Cells were analysed on FACS ARIA II (BD). Data were processed with FACSDiva 8.0 software (BD) and FlowJo v10. FACS plots (Fig. 4a, b) are representative ploidy plots from *n* = 5 wild-type animals and *n* = 3 Axin2-CreERT2;Rosa26-mTmG^{lox} mice, respectively.

Real-time RT-PCR analysis. Liver endothelial cells were FACS sorted as described above. Cells were homogenized using QIAshredder (Qiagen) and total RNA was purified using an RNeasy mini kit (Qiagen) according to the manufacturer's instructions. The total RNA was reverse transcribed using random primers (High Capacity cDNA Reverse Transcription kit, Life Technologies). Gene expression was then assayed by real-time PCR using TaqMan Gene Expression Assays (Applied Biosystems) on an ABI 7900HT real-time PCR system. The following TaqMan probes were used: *Axin2* (Mm00443610), *Wnt1* (Mm01300555), *Wnt2* (Mm00470018), *Wnt9b* (Mm00457102).

RNAscope *in situ* hybridization. Paraffin-embedded liver sections (5 μ m) were processed for RNA *in situ* detection using the RNAscope 2-plex Detection Kit (Chromogenic) according to the manufacturer's instructions (Advanced Cell Diagnostics)⁴³. RNAscope probes used were: *Axin2* (NM 015732, region 330-1287), *GS* (NM 008131, region 103-973), *Wnt2* (NM 023653, region 857-2086), *Wnt9b* (NM 011719, region 727-1616), *PECAM1* (NM 001032378, region 915-1827), *Tbx3* (NM 198052).

Representative figures of *in situ* hybridization of *Tbx3* (Fig. 2d), *Wnt2*, *Wnt9b* and *Pecam1* (Fig. 5a–d) are from *n* = 5 wild-type B6 mice aged 8 weeks.

Representative figures of *in situ* hybridization from VE-cadherin-CreERT2; Wls^{lox} studies (Fig. 6a, b, d, e) are from *n* = 5 mice from each group.

Microscope image acquisition and quantification. All sections were imaged using the Axioplan 2 microscope, the AxioCam MRm (fluorescence) and MRC5 (bright field) cameras and using Axiovision AC software (Release 4.8, Carl Zeiss). Image acquisitions were done at room temperature using $\times 10$ NA 0.3, $\times 20$ NA 0.5, and $\times 40$ NA 0.75 EC Plan-Neofluar objectives. Co-localization images were obtained using confocal microscopy using a Leica SP5 confocal detection system fitted on a Leica DMI6000 inverted microscope equipped with a $\times 20$ NA 0.75 HC PL apochromatic glycerol-immersion objective, and a $\times 40$ NA 1.3 HCX PL apochromatic oil-immersion objective (Leica) and using Leica LAS AF system software. For fluorescence area quantification, tiled images of entire liver sections were acquired using a Zeiss Cell Observer Spinning Disc confocal system on an Axio Observer.Z1 inverted microscope with Zen 2012 software (blue edition). Image acquisitions were done at room temperature using a $\times 20$ NA 0.5 EC Plan-Neofluar objective.

Tiled images were stitched together in Zen 2012 (blue edition) and quantified using ImageJ. For some images, contrast, colour and dynamic range were globally adjusted in Adobe Photoshop (Adobe Systems). Nuclei for EdU-labelled and total cell counts were quantified using ImageJ. Thresholding and watershed transforms were used. Pericentral hepatocytes were identified with either GFP (in Axin2-CreERT2;Rosa26-mTmG^{lox} mice) or GS (in wild-type or VE-cadherin-CreERT2;Wls^{lox} mice). GFP⁺ and GS⁺ cells were quantified manually. Axin2 and GS *in situ* images were quantified with RNAscope SpotStudio software (version 1.0, Advanced Cell Diagnostics)⁴⁴.

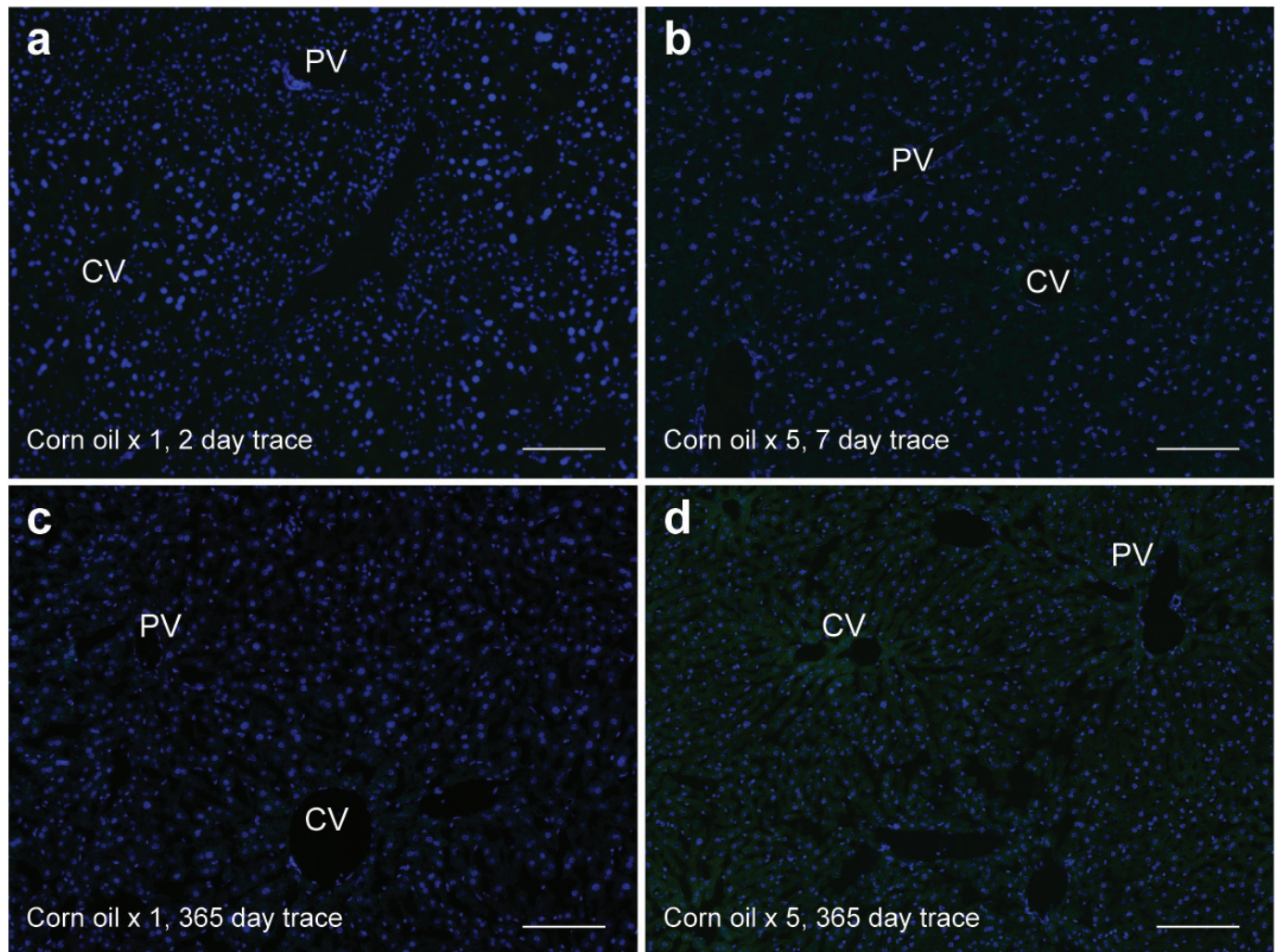
RNA-seq analysis. Using hepatocytes isolated from Axin2-rtTA;TetO-H2B-GFP animals, GFP⁺ and GFP⁻ cells from three animals were sorted using FACS. Cells were lysed in TRIzol (Life Technologies) and treated with chloroform. The aqueous layer was precipitated with ethanol and RNA was isolated with QIAGEN RNeasy Mini kit following the manufacturer's instructions. cDNA barcoded libraries were made using the TruSeq Stranded mRNA Sample Prep Kit (Illumina) following the manufacturer's instructions. Samples were sequenced on an Illumina HiSeq 2000 instrument (three samples per lane, 100-bp paired-end reads) to yield >50 million reads per sample.

Processing and analysis of FASTQ files were performed using Galaxy⁴⁵. A custom Galaxy instance (UC Davis Bioinformatics Core) was run on Amazon AWS. Removal of adaptor contamination and quality trimming were performed using Scythe v1.21 (<https://github.com/ucdavis-bioinformatics/scythe>) and Sickel v1.21 (<https://github.com/ucdavis-bioinformatics/sickel>). TopHat v2.0.11⁴⁶ was

used to align reads to the mouse mm10 assembly, and Cuffdiff v.2.2.2⁴⁷ was used for differential gene expression analysis.

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus⁴⁸ and are accessible through GEO Series accession number GSE68806.

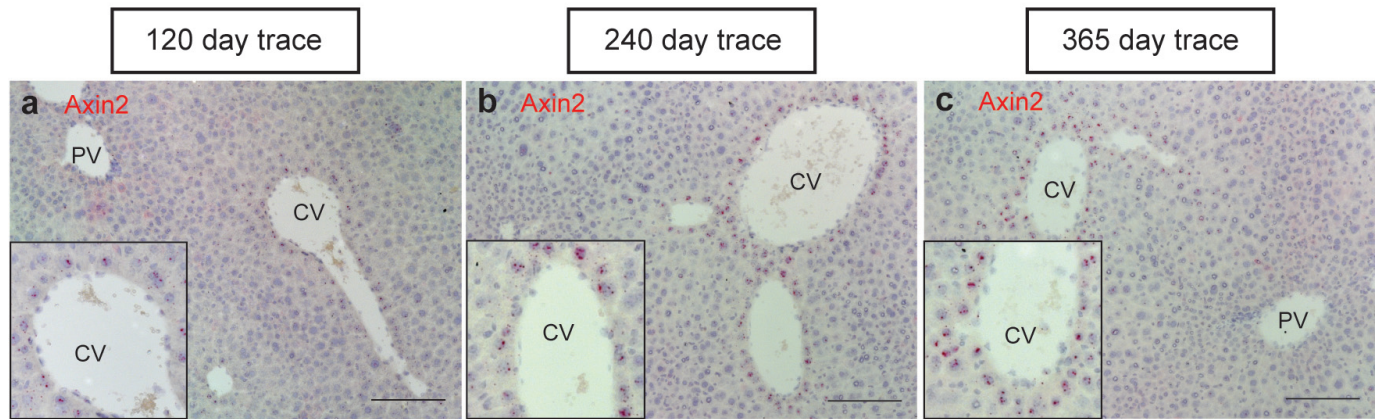
42. Muzumdar, M. D. *et al.* A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).
43. Kreamer, B. L. *et al.* Use of a low-speed, iso-density percoll centrifugation method to increase the viability of isolated rat hepatocyte preparations. *In Vitro Cell. Dev. Biol.* **22**, 201–211 (1986).
44. Wang, F. *et al.* RNAscope: a novel *in situ* RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* **14**, 22–29 (2012).
45. Goecks, J. *et al.* Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86 (2010).
46. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
47. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnol.* **31**, 46–53 (2013).
48. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).



Extended Data Figure 1 | Leakiness in Axin2-CreERT2;Rosa26-mTmG^{lox} mice is not observed in animals injected with corn oil. a, c, No GFP labelling is seen in Axin2-CreERT2;Rosa26-mTmG^{lox} mice after a single dose of corn oil and traced for 2 days (a) or 365 days (c). b, d, No GFP labelling is

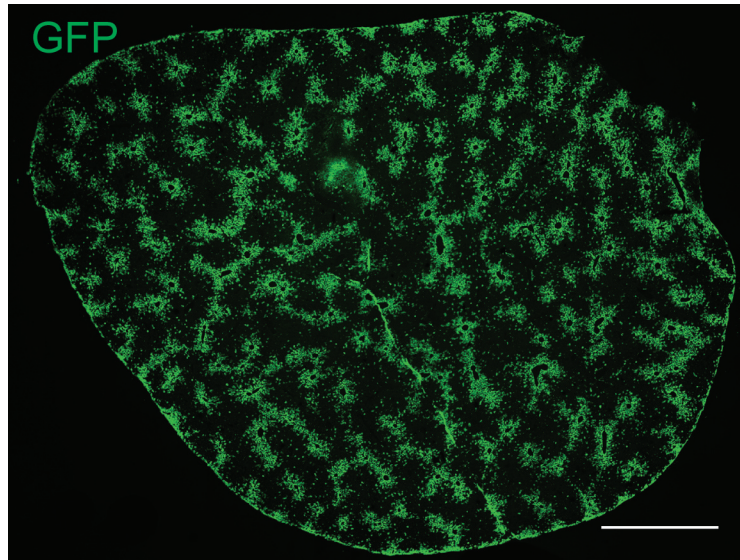
seen after five consecutive daily doses of corn oil and traced for 7 days (b) or 365 days (d). All animals were 8-week-old Axin2-CreERT2;Rosa26-mTmG^{lox} mice. Images are representative images from $n = 5$ mice per condition and time point. CV, central vein; PV, portal vein. Scale bars, 100 μm.

Single tamoxifen injection (4mg/25gm x1)

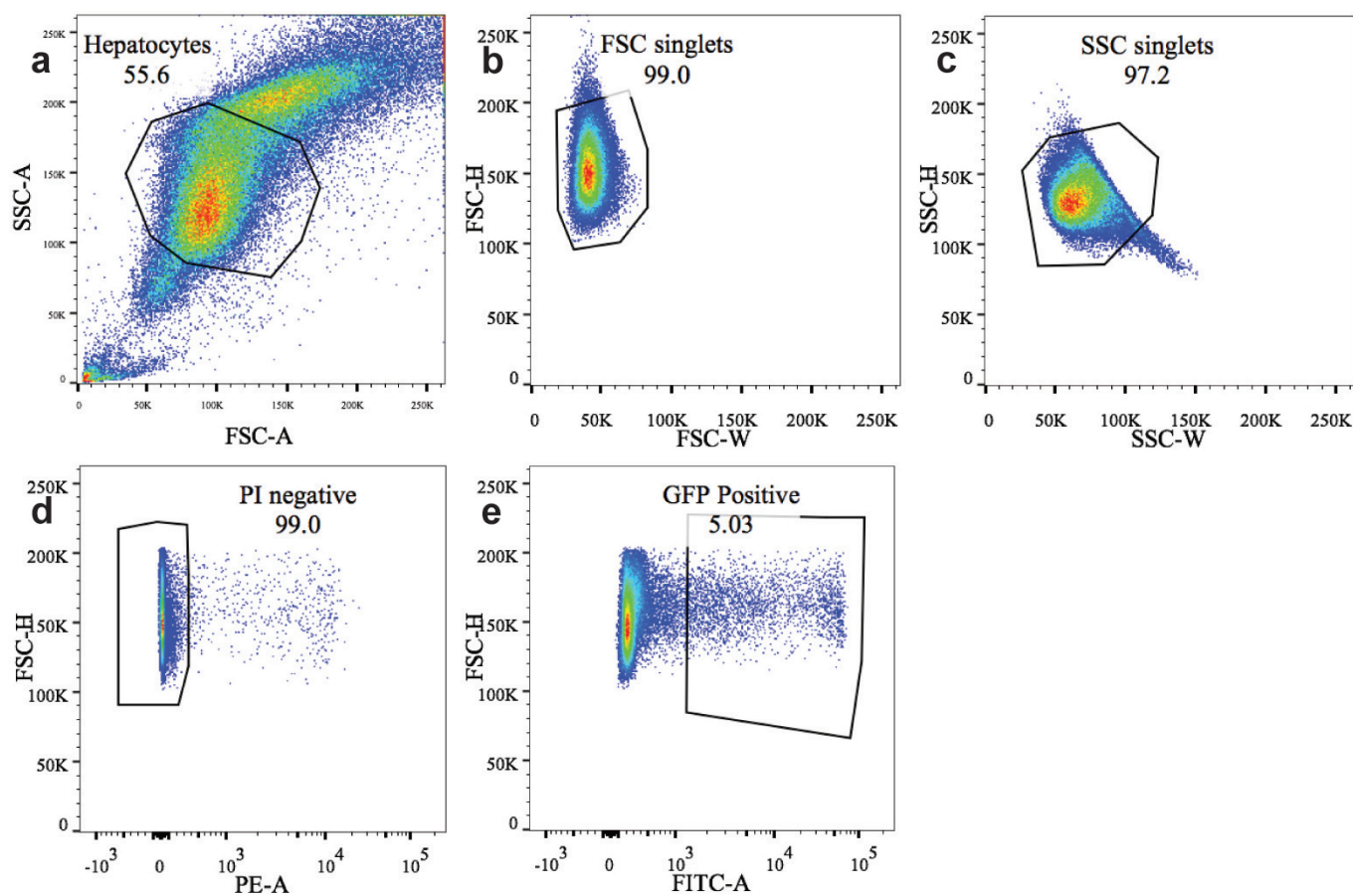


Extended Data Figure 2 | Axin2 expression remains restricted to pericentral cells. a–c, *In situ* hybridization for *Axin2* in 120-day trace (a), 240-day trace (b) and 365-day trace (c) *Axin2*-CreERT2;*Rosa26*-mTmG^{flox} mice.

Representative *in situ* images are from $n = 5$ animals per time point. CV, central vein; PV, portal vein. Scale bars, 100 μ m.

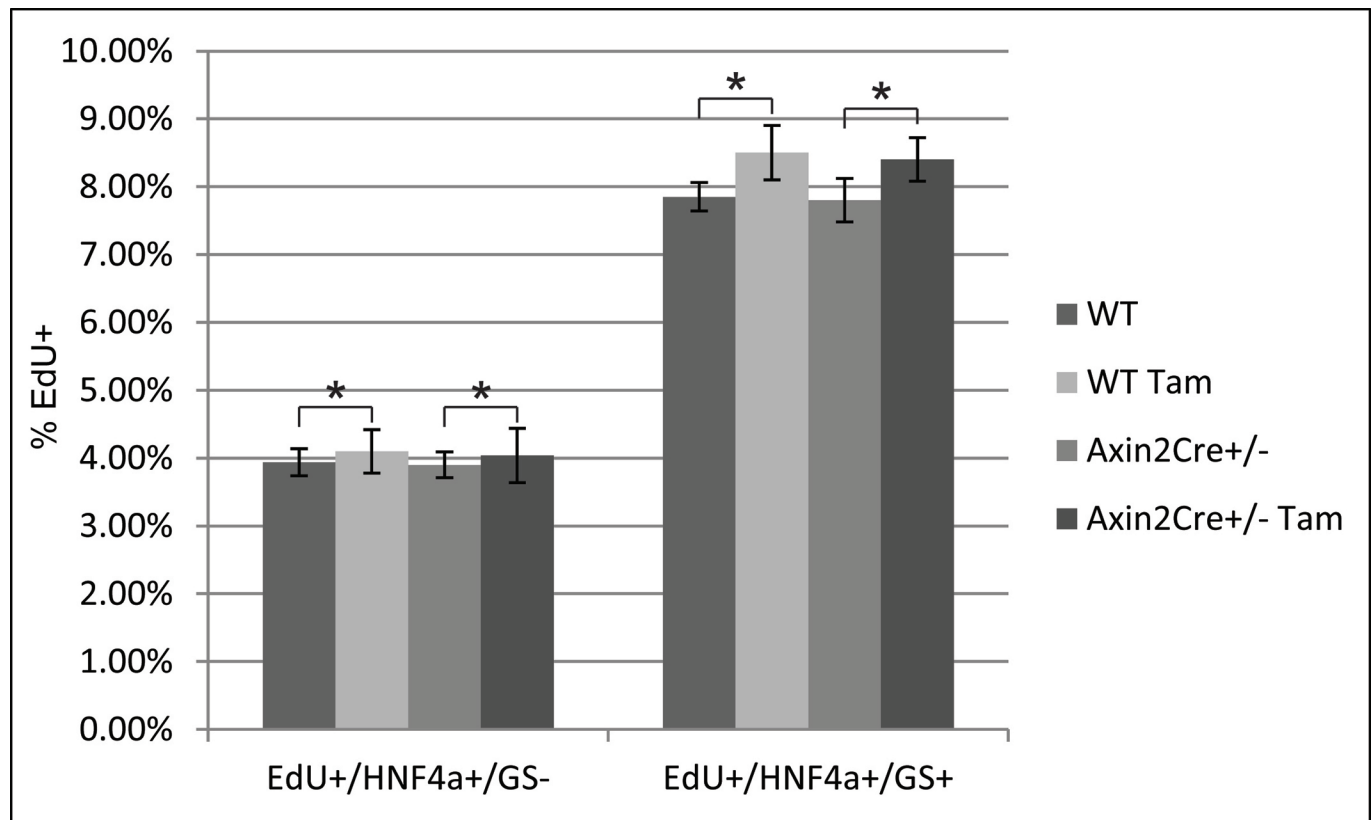


Extended Data Figure 3 | Descendants of Axin2⁺ cells replaced 30% of the area of the liver. Tiled image of entire liver section of a 365-day trace Axin2-CreERT2;Rosa26-mTmG^{fl} mice. Image is representative of $n = 5$ animals at this time point. Scale bar, 2,500 μm .



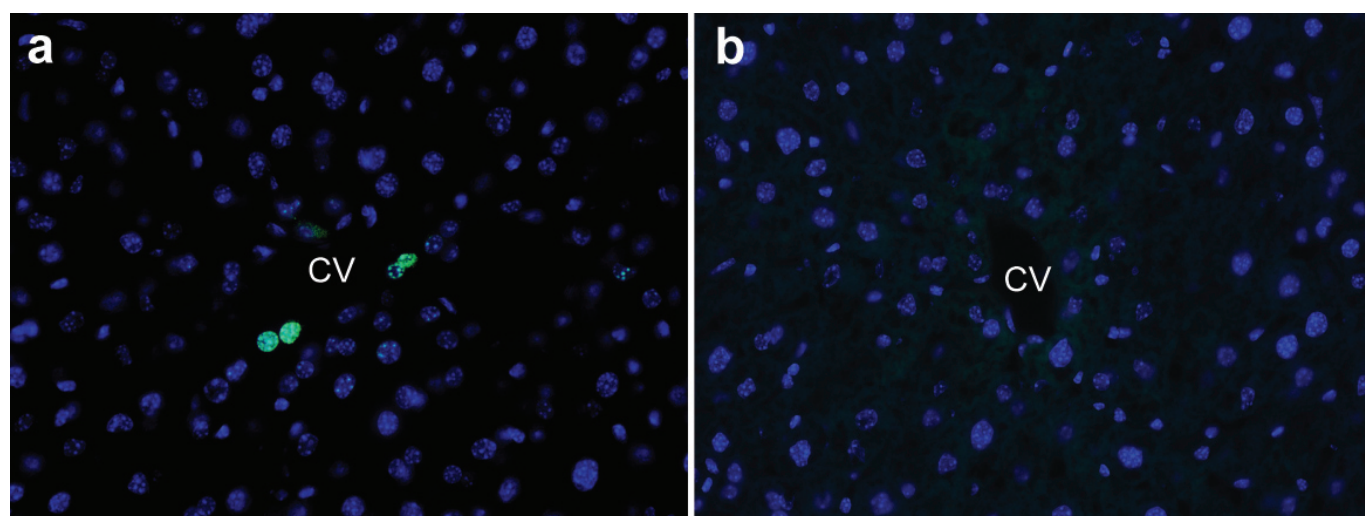
Extended Data Figure 4 | FACS sorting gates for GFP⁺ cells in Axin2-rtTA;TetO-H2B-GFP mice. Eight-week-old Axin2-rtTA;TetO-H2B-GFP mice were labelled with doxycycline for 7 days and chased for various lengths of time. Hepatocytes were enzymatically dispersed and sorted by FACS. **a–c**, Successive gating shows sequential selection of all hepatocytes (**a**), single

cells by forward scatter (**b**), and side scatter (**c**). **d**, Dead cells were excluded by propidium iodide labelling. **e**, GFP-positive cells were gated and either sorted for RNA-seq analysis or further graphed as histograms for GFP intensity analysis (see Fig. 3g).



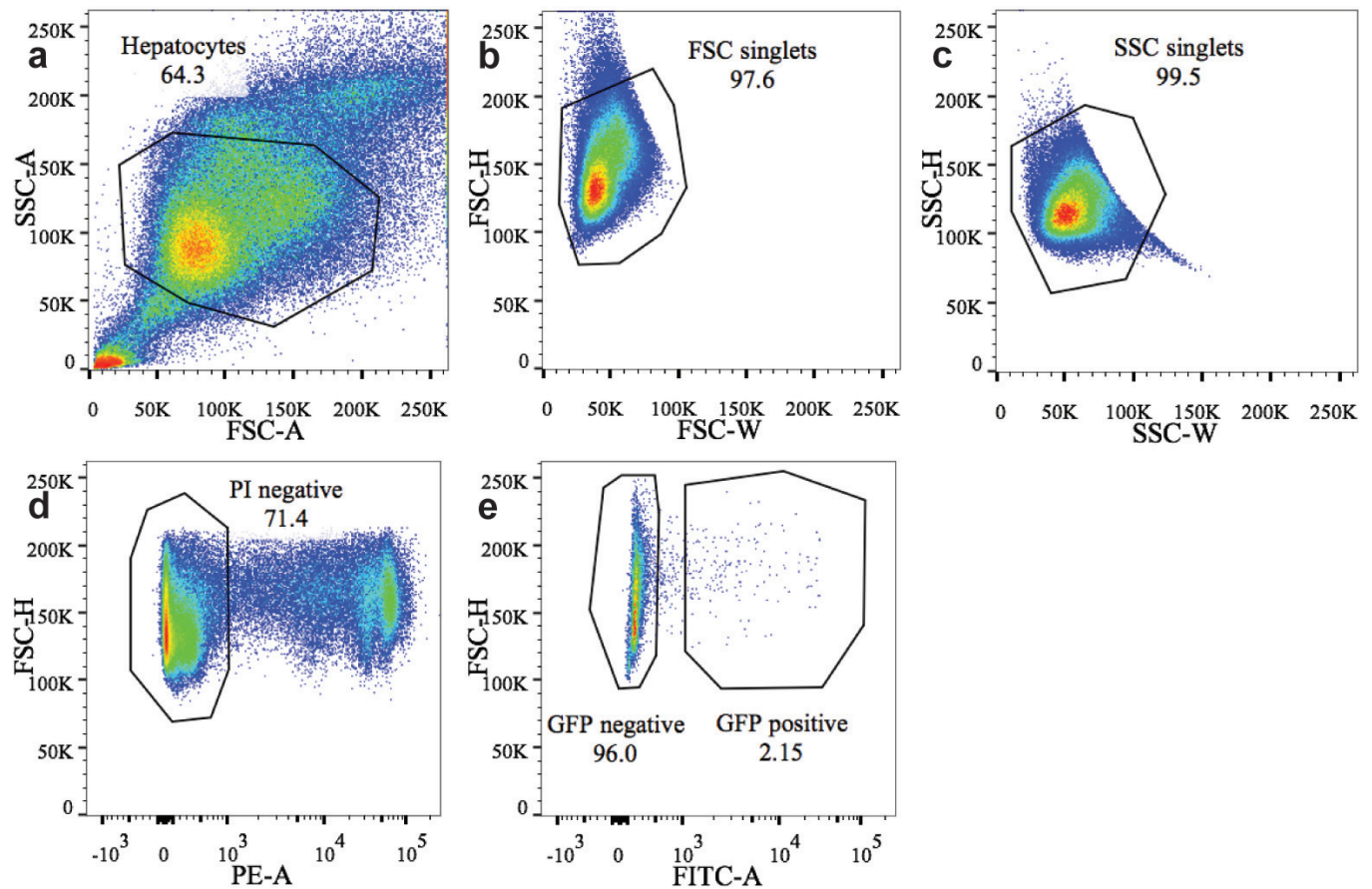
Extended Data Figure 5 | *Axin2* gene dosage and tamoxifen have no effect on pericentral hepatocyte proliferation rate. Wild-type and *Axin2*CreERT2^{+/-} mice were given EdU daily for 7 days. A subset of wild-type and *Axin2*CreERT2^{+/-} mice was given 4 mg of tamoxifen per 25 g body weight daily for 5 days. Pericentral hepatocytes were identified by Hnf4a⁺/GS⁺

staining. All other hepatocytes were identified by Hnf4a⁺/GS⁻ antibody staining. The EdU-positive rates within the two hepatocyte populations as a percentage of total HNF4a⁺ cells were essentially the same regardless of *Axin2* gene dosage or tamoxifen administration. *n* = 5 animals per group. Data represent mean ± s.e.m. **P* > 0.05.



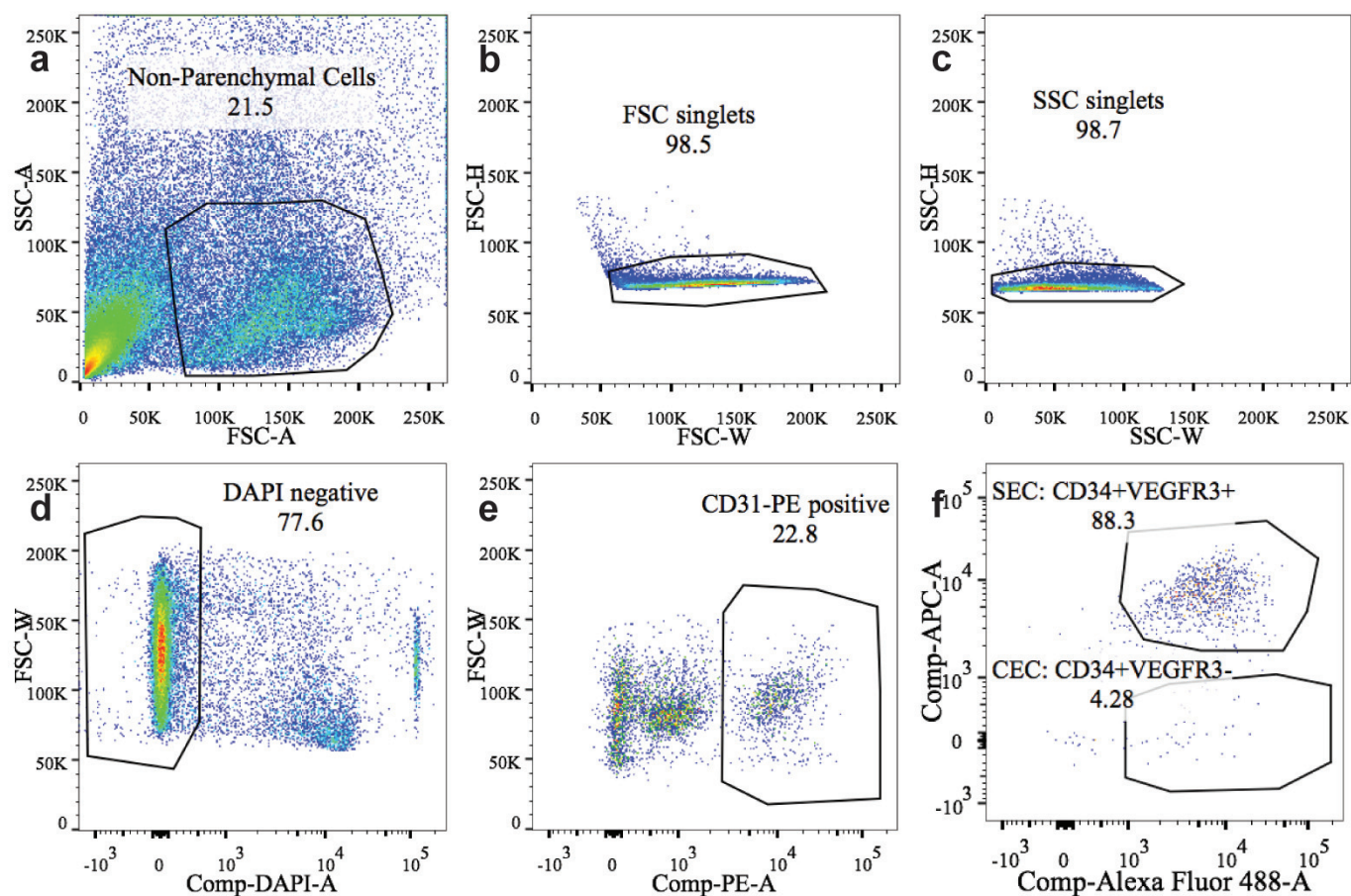
Extended Data Figure 6 | Axin2⁺ hepatocytes proliferate rapidly. Axin2-rtTA;TetO-H2B-GFP mice were given doxycycline for 7 days. **a**, 56 days after cessation of doxycycline, very few GFP⁺ cells are seen around the

central vein. **b**, After 84 days, no GFP⁺ cells are seen. Images are representative of $n = 4$ animals per time point.



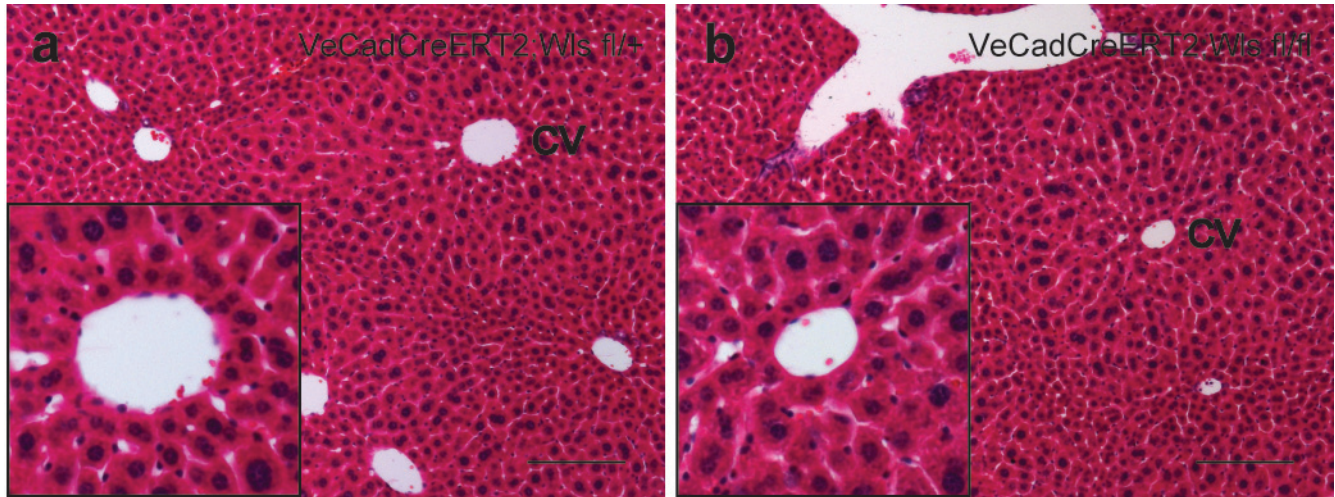
Extended Data Figure 7 | FACS sorting gates for GFP⁺ cells in Axin2-CreERT2;Rosa26-mTmG^{lox} mice for ploidy analysis. Eight-week-old Axin2-CreERT2;Rosa26-mTmG^{lox} mice were labelled with five daily doses of tamoxifen and traced for 7 days. Hepatocytes were enzymatically dispersed

and sorted by FACS. **a–c**, Successive gating show sequential selection of all hepatocytes(**a**), single cells by forward scatter(**b**), and side scatter (**c**). **d**, Dead cells were excluded by propidium iodide labelling. **e**, GFP-positive cells were gated and graphed as histograms for Hoechst staining (see Fig. 4).



Extended Data Figure 8 | FACS sorting gates for endothelial cells. Eight-week-old wild-type C57B6 mice were used for endothelial cell isolation. Livers were enzymatically digested, hepatocytes were removed by centrifugation and nonparenchymal cells were antibody stained and sorted by FACS. **a–c**, Successive gating showed sequential selection of non-parenchymal cells by

size (**a**), single cells by forward scatter (**b**), and side scatter (**c**). **d**, Dead cells were excluded by DAPI labelling. **e**, endothelial cells were identified by CD31-phycoerythrin-positive staining. **f**, Sinusoidal endothelial cells (SEC) were identified as CD34-FITC⁺VEGFR3-APC⁺ while central vein endothelial cells (CEC) were identified as CD34-FITC⁺VEGFR3-APC⁻.



Extended Data Figure 9 | Histology of VE-cadherin-CreERT2;Wls^{flx/flx} animal versus control. **a**, Control (VE-cadherin-CreERT2;Wls^{flx/+}) animals given five daily doses of tamoxifen and traced for 7 days after the last tamoxifen dose. Haematoxylin and eosin staining of the liver shows normal histology.

b, Wls-knockout animals (VE-cadherin-CreERT2;Wls^{flx/flx}) also showed normal liver histology. Images are representative images from $n = 5$ animals per group. Insets show central veins. Scale bars, 100 μm .

Extended Data Table 1 | Partial list of differentially expressed genes in Axin2⁺ vs Axin2⁻ hepatocytes by RNA-seq analysis

Differentially expressed genes:	Log ₂ (fold_change):	q-value:
Pericentral:		
Glutamine synthetase (Glu1)	-2.58912	0.00961319
Leukocyte cell-derived chemotaxin 2 (Lect2)	-1.85582	0.00961319
Axin2	-1.96625	0.00961319
B glycoprotein (Rhbg)	-3.1865	0.00961319
Cytochrome P450 1a2 (Cyp1a2)	-2.02709	0.00961319
T-box transcription factor 3 (Tbx3)	-1.26286	0.0380302
Periportal:		
Asparaginase (Aspg)	1.25545	0.0441798
Glutaminase 2 (Gls2)	1.53395	0.0441798

Cells were isolated from Axin2-rtTA;TetO-H2B-GFP mice after labelling with doxycycline. Genes preferentially expressed in Axin2⁺ hepatocytes are indicated by negative log₂ fold changes and genes preferentially expressed in Axin2⁻ hepatocytes are indicated by positive log₂ fold changes. q value (false-discovery-rate-adjusted *P* value) <0.05 used to determine significance.

Structure of the eukaryotic MCM complex at 3.8 Å

Ningning Li^{1*}, Yuanliang Zhai^{2,3*}, Yixiao Zhang¹, Wanqiu Li¹, Maojun Yang¹, Jianlin Lei¹, Bik-Kwoon Tye^{2,4} & Ning Gao¹

DNA replication in eukaryotes is strictly regulated by several mechanisms. A central step in this replication is the assembly of the heterohexameric minichromosome maintenance (MCM2–7) helicase complex at replication origins during G1 phase as an inactive double hexamer. Here, using cryo-electron microscopy, we report a near-atomic structure of the MCM2–7 double hexamer purified from yeast G1 chromatin. Our structure shows that two single hexamers, arranged in a tilted and twisted fashion through interdigitated amino-terminal domain interactions, form a kinked central channel. Four constricted rings consisting of conserved interior β -hairpins from the two single hexamers create a narrow passageway that tightly fits duplex DNA. This narrow passageway, reinforced by the offset of the two single hexamers at the double hexamer interface, is flanked by two pairs of gate-forming subunits, MCM2 and MCM5. These unusual features of the twisted and tilted single hexamers suggest a concerted mechanism for the melting of origin DNA that requires structural deformation of the intervening DNA.

For DNA to be replicated, two strands of the duplex DNA must be separated so that each can serve as a template for the synthesis of daughter strands. In both prokaryotes and eukaryotes, DNA unwinding is carried out by specialized helicases that encircle and translocate along one of the DNA strands¹. However, the mechanisms for the initial melting or unwinding of origin DNA are markedly different^{1,2}. In bacteria, the origin recognition protein DnaA initiates origin melting and then recruits hexameric helicase DnaB³ to unwind DNA by translocation on the lagging strand in the 5′–3′ direction². By contrast, in eukaryotes, the origin recognition complex first binds replication origins without effecting initial origin melting, and then loads two MCM2–7 (ref. 4) single hexamers with the help of the DNA replication proteins Cdc6 and Cdt1 onto double-stranded origin DNA to form a double hexamer^{5,6}. This inactive assembly of proteins is known as the pre-replicative complex. Subsequent activation of MCM2–7 complex takes place in the S phase, and requires several factors and cell-cycle-specific kinases^{7–11}, resulting in the formation of an active replicative helicase, the Cdc45–MCM2–7–GINS (CMG) complex¹². CMG translocates along the leading strand in a 3′–5′ direction to unwind duplex DNA (steric exclusion model)^{13–17}. However, how origin DNA is melted before active replication elongation is unknown. This process probably requires the reconfiguration of MCM2–7 helicase, a complex molecular motor that has defied high-resolution structural analysis for decades. At present, much of the mechanistic insights came from low-resolution structures of the MCM2–7 complex in functional forms from different species^{5,16,18–22}, as well as crystal structures of simpler archaeal versions, in non-functional oligomers^{23–25}, truncations^{26–29} or a chimaeric hexamer³⁰.

In this study, we purified the endogenous MCM2–7 double hexamer from G1 chromatin of budding yeast (Extended Data Fig. 1a–c), and determined its cryo-EM structure at an overall resolution of 3.8 Å (gold-standard Fourier shell correlation 0.143 criteria) (Extended Data Fig. 1k). Except for peripheral regions, the core of the map is better than 3.5 Å (Extended Data Fig. 1j), which enabled atomic model building for ~80% of sequences of this 1.2-megadalton (MDa) complex. Our structure reveals rich details for the organiza-

tion of this large complex, and informs many functional aspects of this replicative helicase, particularly in the initial origin melting.

Overall structure and domain organization

A first glimpse of the structure is the tilted arrangement of two single hexamers, with a 14° wedge in between (Fig. 1a–c), a feature already noticed from low-resolution data of the MCM2–7 double hexamer^{5,18} and the SV40 large tumour antigen³¹. The two single hexamers also have a twisted arrangement (Fig. 1a, side panel), resulting in the misalignment of two hexamer axes. The quality of the density map allowed an independent assignment of six subunits, being 2–6–4–7–3–5 (viewed from the carboxy-terminal domain (CTD) ring) (Fig. 1d), consistent with the well-established model^{21,32,33}. Notably, when viewed from the single hexamer axis, the gravity centres of three major structural components—NTD-A (A subdomain of N-terminal domain (NTD)), oligonucleotide/oligosaccharide-binding (OB)-fold (C subdomain of NTD), and CTD—fall onto three eccentric circles (Fig. 1d). While the circles of NTD-As and OBs are nearly concentric, the CTD circle exhibits apparent rotational and translational offsets, indicating a relative shift and twist between the NTD and CTD rings within the single hexamer. Also, the NTD-As and OBs for each subunit are nearly vertically arranged (as indicated by their centres falling along on the same radial lines), with slight rotations in opposite directions for MCM4 and MCM5. Notably, the CTDs of all six subunits have left-handed twists to varying extents (Fig. 1d, f) with respect to their OBs and NTDs. Furthermore, the distances between neighbouring CTDs are different (Fig. 1e), showing a 4–Å difference between tightly (4:7, 7:3 and 5:2) and loosely (6:4, 3:5 and 2:6) packed groups. While six OBs form a plane perpendicular to the hexamer axis, the CTDs and NTD-As display marked axial variations (Fig. 1f, g).

Inter-hexamer interface

The head-to-head stacking of the two hexamers is largely mediated by their zinc-fingers (ZFs; B subdomain of NTD), as expected from

¹Ministry of Education Key Laboratory of Protein Sciences, Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China. ²Division of Life Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China. ³Institute for Advanced Study, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China.

⁴Department of Molecular Biology and Genetics, College of Agriculture and Life Sciences, Cornell University, Ithaca, New York 14853, USA.

*These authors contributed equally to this work.

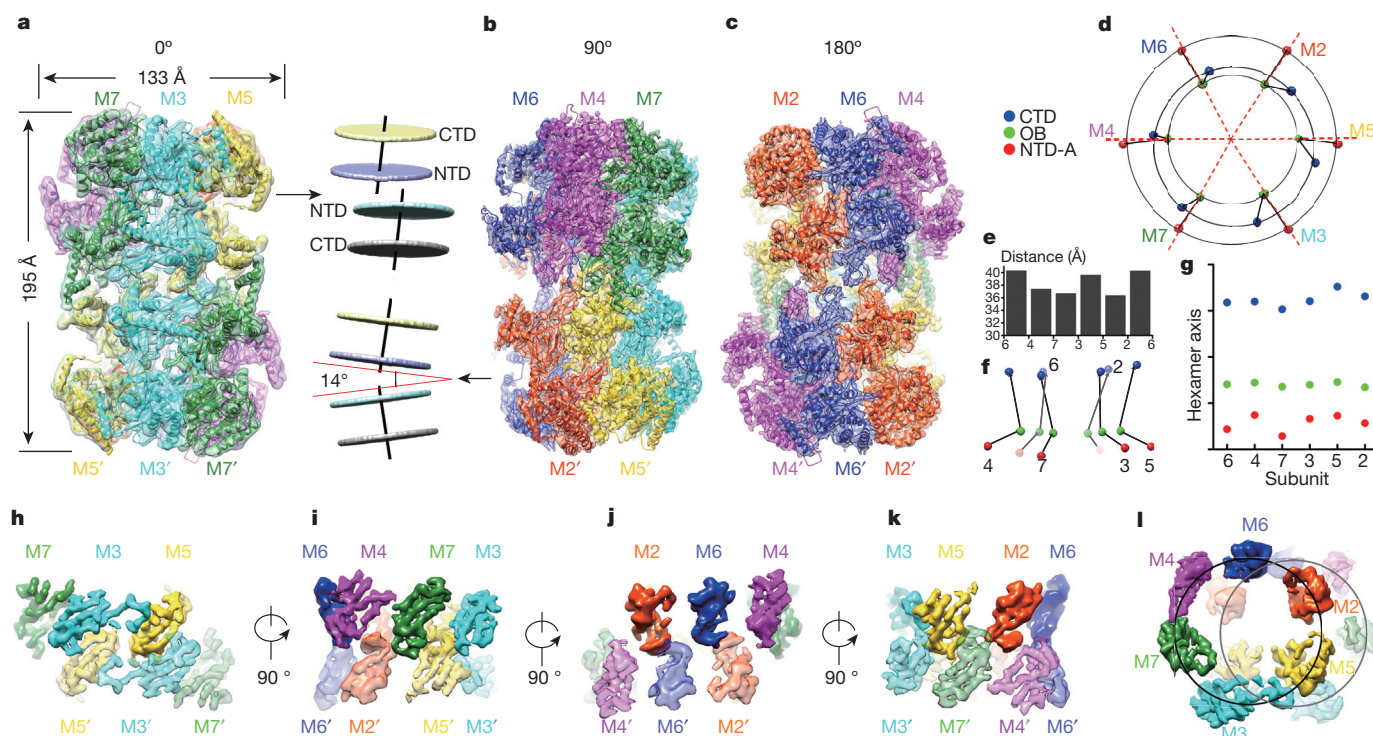


Figure 1 | Overall structure and domain organization of the MCM2-7 double hexamer. **a–c**, Side-views of the cryo-EM density map superimposed with the atomic model. Unsharpened map (**a**) is displayed from the two-fold axis, and sharpened map (**b**, **c**) displayed with indicated rotations relative to **a** along the cylinder axis. The side panels of **a** and **b** illustrate the tilted

arrangement of the two single hexamers. **d**, **f**, Top (**d**) and side (**f**) views of the organization of the indicated structural domains. Small coloured balls denote gravity centres (see Methods) of these domains. **e**, Distances between the gravity centres of adjacent CTDs. **g**, Radial projection of **f**. **h–l**, Side and top views of the segmented maps (sharpened) of ZFs at the hexamer interface.

previous studies^{26,34,35}. Notably, consistent with sequence analysis³⁶ (Extended Data Fig. 2), the ZF of MCM3 is a degenerate version without zinc binding (Extended Data Fig. 3a–c). Twelve ZFs arrange into two stacked rings at the interface (Fig. 1h–l), with an apparent centre shift (Fig. 1l and Supplementary Video 1). Inter-ZF interactions are versatile, displaying completely different patterns at opposite sides of the wedged interface (Fig. 1b, subpanel). Although ZFs are more horizontally arranged at the thin 3-5-7 edge (Fig. 1h), they are nearly vertical at the thick 2-4-6 edge (Fig. 1j). ZF interactions are largely from their polar residues, dominated by two pairs of ZF5:ZF3' (Fig. 1h) and two pairs of ZF6:ZF2' (Fig. 1j), as measured in buried surfaces (Extended Data Table 1a). Different ZF orientations at the hexamer interface perfectly explain the observed tilt and twist between the two single hexamers, because this unique arrangement would enable comprehensive close contacts for both edges and leads to the stabilization of the double hexamer.

Eukaryotic MCM proteins distinguish themselves from archaeal counterparts by many subunit-specific sequence extensions at their N and C termini (NTE and CTE, respectively) and insertions within functional domains. Both MCM4 and MCM6 have a very long linker between their OBs and CTDs (Extended Data Fig. 4a), which could be the underlying basis of the observed twist between NTD and CTD rings in the single hexamer (Fig. 1d). Notably, many sequence insertions and extensions also markedly contribute to the double hexamer stabilization (Fig. 2). For example, an insertion located on the β -turn of the OB from MCM6 (Extended Data Fig. 4h) interacts with the ZF of MCM2 on the other hexamer (Fig. 2e). The most unique inter-hexamer interactions involve MCM3, MCM5 and MCM7. Compared with archaeal MCMs, they have longer sequences at their N termini (Extended Data Fig. 2), which form extended strands or loops (Extended Data Fig. 4e, g and i). MCM7 also has a long insertion (~70 residues) at its NTD-A (N-terminal insertion (NTI)) (Extended Data Fig. 4a), folding into a helix–turn–helix motif (Extended Data

Fig. 4i). The N terminus of MCM5 extends into the space between ZFs of MCM3 and MCM7 from the other hexamer, and forms interactions with β -strands of both ZFs (Fig. 2f and Supplementary Video 2). Furthermore, the long NTI of MCM7 extends towards the opposite MCM5 and interacts with its NTD-A (Fig. 2c, g). The N terminus of MCM7 also interacts with the N terminus of MCM3 from the other hexamer (Fig. 2h). On the basis of the calculated buried surfaces for the above interfaces (Fig. 2e–h), the contribution of NTIs and NTEs to the double hexamer stability is even greater than the ZF interactions (Extended Data Table 1a). Importantly, most insertions and extensions involved in inter-hexamer stabilization (Fig. 2) are conserved in higher eukaryotes, suggesting a universal importance of these eukaryotic-specific sequences.

MCM2, 4 and 6 have very long NTEs, which are targets of cellular signalling kinases^{7–9,11}. These NTEs are highly disordered in our structure, and their involvement in inter-hexamer interaction is unknown.

Intersubunit interaction

The intersubunit interactions are very similar, and can be categorized into three tiers based on their axial locations (Extended Data Fig. 5a–c). The first one is between two contacting CTDs (ATPase domains), largely composed of hydrophobic interactions, as exemplified by a tight stacking between two surface-exposed helices from two respective CTDs (Extended Data Fig. 5a). The second tier, on the neck region of the hexamer, involves four conserved hairpins or loops from two adjacent subunits, including allosteric communication loop (ACL) and helix-2-insert (H2I) of the first subunit, and H2I and presensor 1 β -hairpin (PS1-HP) of the flanking second subunit (Extended Data Fig. 5b). Atomic interactions at the neck interfaces are versatile, involving different residues from these loops. However, a large proportion of them are polar residues, indicating electrostatic or hydrogen-bonding interactions dominate these interfaces. The third tier, contributed by the ZF of one subunit and two loops from the OB of

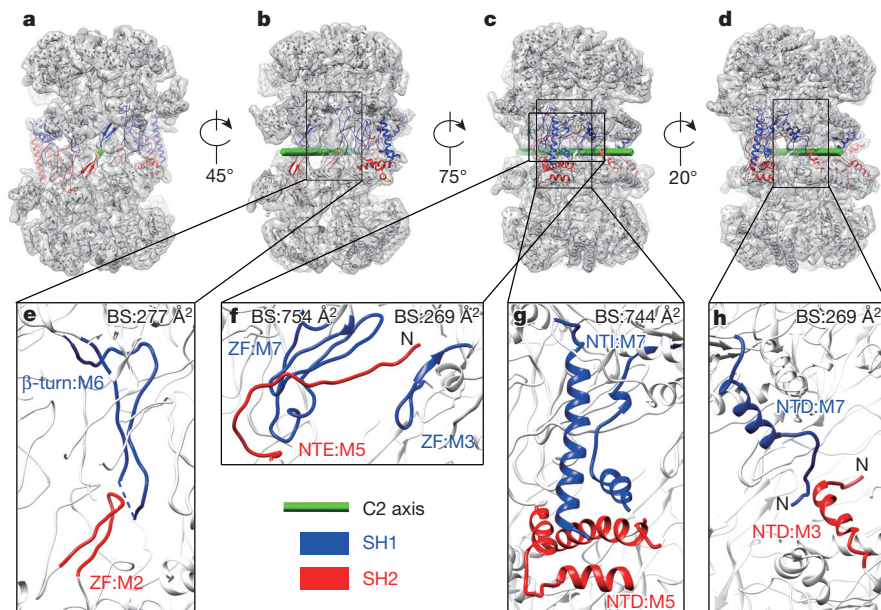


Figure 2 | Inter-hexamer interactions contributed by NTEs and NTIs. a–d, Side views of the MCM2–7 double hexamer, with indicated rotations around the cylinder axis. Atomic structure is superimposed with the unsharpened map. The sequence elements involved in inter-hexamer interactions are highlighted in blue and red representing single hexamer 1 (SH1) and SH2, respectively. e–h, Zoomed-in views of the boxed regions from b–d. Buried areas (Å²) of these interfaces in e–h are labelled. BS, buried surface.

a flanking subunit (Extended Data Fig. 5c), is largely composed of hydrophobic interactions between respective β -strands or loops (for example, see Extended Data Fig. 6d). Perturbation of this interface by an MCM4 mutation (Phe391Ile) causes pre-replicative complex assembly defects in yeast and mammary carcinoma in mouse³⁷. In addition, a mutation on MCM5 NTD-A (Phe83Leu) that results in Dbf4-dependent kinase (DDK)-independent activation³⁸ is close to this interface.

Intersubunit interactions are further enhanced by the NTIs of MCM3, 5, 6 and 7, which contact the NTD-As of their neighbouring MCM7, 3, 2 and 4, respectively (Extended Data Fig. 5d–i). Compared with other pairs, the interface of MCM5–MCM2 is without NTI involvement, a feature that may facilitate the gap opening observed between them during hexamer loading and activation^{16,19–21}. At lower contour levels, four CTEs containing the winged-helix DNA binding motif could be identified for MCM4, 5, 6 and 7 (Extended Data Fig. 3d, e). The flexibility of these winged-helix-containing CTEs suggests that they are not involved in intersubunit interaction, contrasting the role of winged-helix motifs in the origin recognition complex structure³⁹.

The buried surfaces of the six subunit interfaces are sharply different (Extended Data Table 1a), with the smallest at 2:6, rather than at the gate-forming 2:5 interface. The weak 2:6 interface gives rise to a unique side channel (13 Å), enough to accommodate single-stranded DNA (ssDNA) (Extended Data Fig. 7 and Supplementary Video 1), in contrast to archaeal MCM structure with six side channels²³. A ssDNA extrusion model has been proposed for the function of side channels in DNA unwinding^{23,35,36}. However, unwinding studies^{13–15,17} generally conflict with this model. A definitive function for this unique 2:6 side channel remains to be examined.

ATPase active centres

Remarkable conformational differences lie at the six ATPase centres of the CTD ring. Comparisons of them indicate that two active centres, 2:6 and 5:2, are apparent outliers. Their ATP-binding pockets are less compact, with sensor elements (sensors 2 and 3, and arginine finger) in MCM2 and MCM6, respectively, considerably shifted away from nucleotides, and the displacements are as large as 4–5 Å measured by C α atoms of sensor 3 residues (Fig. 3a, b). Further analysis was done by comparisons of representative centres from the compact and loose groups with an active ATPase centre from papillomavirus E1 crystal structure⁴⁰ or an inactive one from an archaeal MCM

structure³⁰. Indeed, while the conformational differences of the four compact centres relative to that of E1 are small (for example, dimers of 7:3 and 4:7; Extended Data Fig. 8h, i), the loose ones display sharp differences from that of E1 (Fig. 3c, d). Moreover, the nucleotide occupancies at the centres of 3:5 and 6:4 dimers are comparatively low (Extended Data Fig. 9), consistent with their reported nearly null ATPase activities³³. On the basis of the active centre arrangement and nucleotide occupancy, it appears that only dimers of 7:3 and 4:7 are active. This observation agrees with the extremely low ATPase activity observed in MCM2–7 as a double hexamer¹⁸, and the reported activity of 7:3 dimer comparable to that of the whole MCM2–7 complex³³.

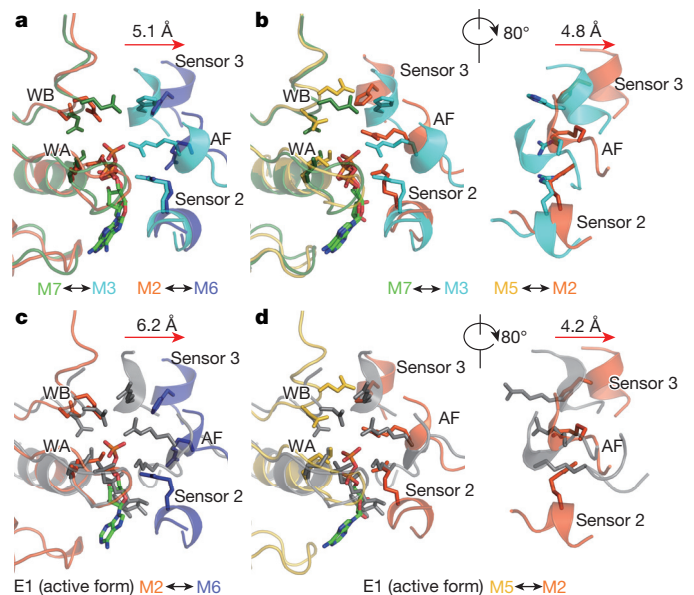


Figure 3 | Inactive ATPase centres of 2:6 and 5:2 dimers. a, b, Comparison of the ATP-binding pockets from 2:6 and 5:2 dimers with that from 7:3 dimer. Dimers (*cis:trans*) are presented with *cis* and *trans* ATPase elements on the left and right, respectively. c, d, Same as a and b, but the ATPase centres of 2:6 and 5:2 dimers are compared with that of the hexameric E1 helicase⁴⁰. Motifs of Walker A (WA), Walker B (WB), arginine finger (AF), sensor 2 and sensor 3 are displayed in stick model. The directions and distances of sensor 3 movements are marked. All alignments were done using the Walker A and B motifs as a reference.

Previously, individual active centres were proposed to have distinct roles in regulating helicase activities^{36,41,42}. Supported by our data, allosteric regulation of these ATPase centres orchestrated by the orientation changes between adjacent CTDs, might be the basis for factor-dependent control of helicase activities during different replication stages.

Axial displacement of interior hairpin loops

As in many hexameric AAA+ machineries, the central-pored chamber of MCM2–7 complex is decorated with layers of hairpin loops. For archaeal MCM, four layers of conserved loops essential for DNA binding and/or unwinding have been described^{35,43}, with two of them located innermost (Fig. 4a). The first one, composed of six H2Is, was previously shown to undergo axial movement depending on the nucleotide-binding states of the ATPase domains⁴⁴. The other, composed of β -turn motifs of the six OBs, was shown to coordinate the binding of ssDNA to the MCM–ssDNA binding motifs on the channel surface of OBs in the crystal structure of an archaeal MCM NTD homohexamer²⁷. In the MCM2–7 complex, these two layers of loops are placed in axially staggered positions (Fig. 4b, c), and particularly, six H2Is roughly display a helical trajectory (Fig. 4c). An alignment with the ssDNA-bound archaeal MCM hexamer²⁷ precisely placed ssDNA between these two layers of loops (Fig. 4b, c). In addition, when a double-stranded DNA (dsDNA) is placed in the channel, the H2Is show very close contact with it, capable of inserting their terminal loops into its major or minor grooves consecutively (Fig. 4d). At a very low threshold, an extra piece of fragmented density, which might be the residual dsDNA, could be identified within the channel at the H2I ring, but its sub-stoichiometric occupancy prevented positive identification and further analysis. Nevertheless, the snug fitting of the helically arranged H2Is and dsDNA suggests that H2I might be

involved in the initial melting of origin DNA. These observations, together with repeated reports of the axial displacement of interior loops of AAA+ hexameric machines^{40,45}, suggest that the MCM2–7 complex uses a conserved mechanism involving cycles of ATP binding and hydrolysis to control the axial positions of the interior loops to facilitate DNA translocation and unwinding.

Central channel and model of initial origin melting

The diameter of the central channel in MCM2–7 single hexamer is not uniform (Fig. 5a, b), about 30 Å at the C-terminal end and 40 Å at the N-terminal end, but with two constriction sites (~ 25 Å) at H2Is and β -turn motifs that are just wide enough to accommodate dsDNA (Fig. 5c, d). However, owing to the twisted stacking between two ZF rings in the double hexamer (Fig. 5f), the channel is partially blocked at the hexamer interface by the ZF rings, splitting the wide channel at the double hexamer interface into a main central channel and two minor channels (Fig. 5f and Supplementary Video 1). The overlapping central channel is just about the size of dsDNA (Fig. 5f, g), while the minor channels are not wide enough for the passage of dsDNA (Fig. 5h) but accessible from the outside. Notably, gate-forming subunits MCM2 and MCM5 participate in the formation of both channels. The overlapping central channel is delineated by ZFs from two MCM2–MCM5 dimers and a vertically arranged MCM6 dimer (Fig. 5g), and the minor channel involves ZFs from MCM2 and 5 of one single hexamer, and MCM3 and 7 of the other single hexamer (Fig. 5h).

The structure of an already constricted central channel of the single hexamer that opens to a larger channel at the NTD only to be occluded by the offset of the ZF rings at the double hexamer interface invites

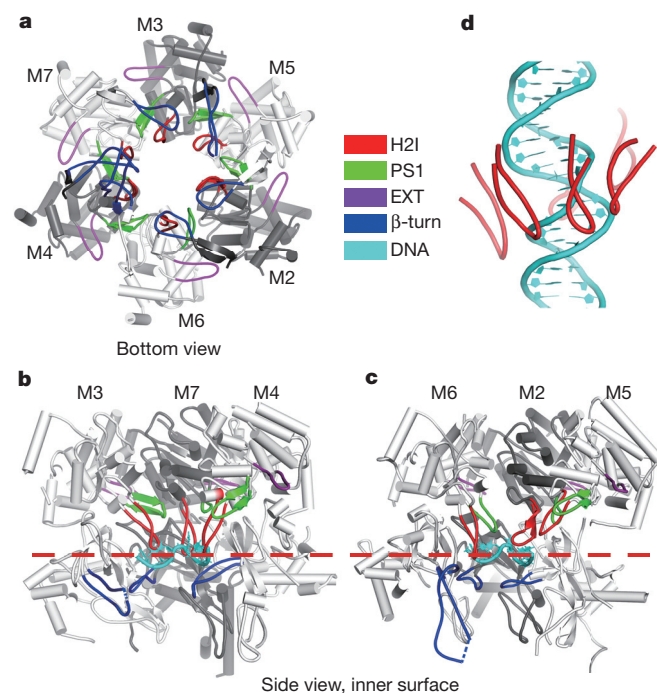


Figure 4 | Spatial distribution of conserved hairpins in the single hexamer. **a**, Bottom view (from the hexamer interface) of the MCM2–7 hexamer. The six MCM subunits are shown in alternating grey scales, with four hairpins highlighted in indicated colours. All four hairpins are at the subunit interface, and extend towards the adjacent subunit. **b**, **c**, Side views of the inner surface of the hexamer, docked with a model ssDNA from an archaeal MCM structure (PDB accession code 4POG)²⁷. **d**, Contacts between H2Is and dsDNA illustrated through fitting a dsDNA fragment into the central channel of the hexamer.

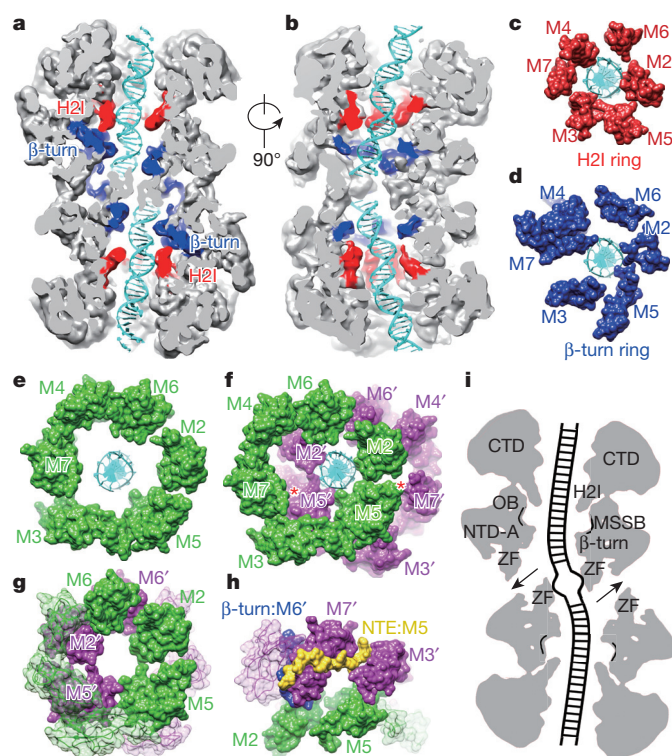


Figure 5 | Central channel and its implication in origin melting. **a**, **b**, Cut-away views of the density map (unsharpened) with two dsDNA fragments fitted in the central channel. **c–f**, Surface representation (top view) of the atomic models of the six H2Is (**c**), β -turns (**d**), ZFs from one single hexamer, and ZFs from both single hexamers (**f**). Red asterisks mark the positions of two minor channels. **g**, **h**, Double hexamer interface centred for the views of the narrowed central channel (**g**), and minor channel (**h**). ZFs not part of the constricted central channel are in high transparency. **i**, Model for initial origin melting (see text).

speculations for functions. First, the kink in the central channel created by the offset of the two single hexamer rings will probably cause deformation of trapped duplex DNA (Fig. 5i) to create a nucleation centre for DNA melting. Second, the tight grip of the duplex DNA on either ends by the helically positioned H2Is serves to hold the kinked DNA in place such that a slight left-handed rotation between the two single hexamers, as previously proposed¹⁸, could further deform the origin DNA at the nucleation point. Third, possible relative rotation between the NTD and CTD rings within single hexamers upon helicase activation, might further lower the activation energy of DNA melting. We envision that initial melting involves allosteric conformational changes, in combination with dsDNA translocation in opposite directions by the coupled single hexamers³². The dsDNA being pumped into the central channel provides the slack necessary for strand separation. This initial melting step requires the activation of the MCM2–7 helicase activity most likely by DDK phosphorylation and binding of Cdc45 and GINS¹². Recent studies showed that DDK phosphorylation of the NTEs of MCM2–7 does not cause double hexamer separation^{18,46}, but promotes MCM2–MCM5 gate opening⁴⁷. Opening of the MCM2–MCM5 gates at this point would merge central and minor channels, creating an expanded N-terminal chamber for strand separation. The ssDNA looping out through this chamber would be accessible to replication factors lurking nearby (Fig. 5i). Further strand separation towards the CTD ring may be facilitated by the interior β -hairpin loops and the MCM–ssDNA binding motifs²⁷ on the inner surface of OBs.

This structure-informed hypothesis on the initial origin melting is in accordance with previous data. First, many factors required for helicase activation, such as Sld2, Sld3, Cdc45 and Mcm10, have well-defined ssDNA binding activity^{11,16,32,48–50}. Second, a similar replicative helicase SV40 large tumour antigen initiates origin melting as a dsDNA pump^{32,35}, and conformational rearrangements of the two single hexamers were observed during this process³¹. Our structure suggests that, in addition to its role in processive fork unwinding, MCM2–7 is also actively involved in origin DNA melting. In transitioning from the initial origin melting state to the fork unwinding state, MCM2–7 essentially translocates first on dsDNA (dsDNA pump) and then along ssDNA (steric exclusion).

In summary, the fine structural details provided in this work will serve as a rich source of information for designing and interpreting biochemical studies aimed at dissecting the mechanistic functions of the MCM2–7 complex. In particular, it will provide a framework for future study of the eukaryotic-specific assembly, activation and regulation of this helicase family.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 March 2015; accepted 25 June 2015.

Published online 29 July 2015.

- O'Donnell, M., Langston, L. & Stillman, B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb. Perspect. Biol.* **5**, a010108 (2013).
- Costa, A., Hood, I. V. & Berger, J. M. Mechanisms for initiating cellular DNA replication. *Annu. Rev. Biochem.* **82**, 25–54 (2013).
- Duderstadt, K. E. & Berger, J. M. A structural framework for replication origin opening by AAA+ initiation factors. *Curr. Opin. Struct. Biol.* **23**, 144–153 (2013).
- Tye, B. K. MCM proteins in DNA replication. *Annu. Rev. Biochem.* **68**, 649–686 (1999).
- Remus, D. *et al.* Concerted loading of Mcm2–7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139**, 719–730 (2009).
- Evrin, C. *et al.* A double-hexameric MCM2–7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proc. Natl Acad. Sci. USA* **106**, 20240–20245 (2009).
- Siddiqui, K., On, K. F. & Diffley, J. F. Regulating DNA replication in eukarya. *Cold Spring Harb. Perspect. Biol.* **5**, a012930 (2013).
- Heller, R. C. *et al.* Eukaryotic origin-dependent DNA replication *in vitro* reveals sequential action of DDK and S-CDK kinases. *Cell* **146**, 80–91 (2011).
- Yeeles, J. T., Deegan, T. D., Janska, A., Early, A. & Diffley, J. F. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**, 431–435 (2015).

- Tanaka, S. & Araki, H. Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harb. Perspect. Biol.* **5**, a01037 (2013).
- Tognetti, S., Riera, A. & Speck, C. Switch on the engine: how the eukaryotic replicative helicase MCM2–7 becomes activated. *Chromosoma* **124**, 13–26 (2015).
- Ilves, I., Petojevic, T., Pesavento, J. J. & Botchan, M. R. Activation of the MCM2–7 helicase by association with Cdc45 and GINS proteins. *Mol. Cell* **37**, 247–258 (2010).
- Fu, Y. V. *et al.* Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* **146**, 931–941 (2011).
- Rothenberg, E., Trakselis, M. A., Bell, S. D. & Ha, T. MCM forked substrate specificity involves dynamic interaction with the 5'-tail. *J. Biol. Chem.* **282**, 34229–34234 (2007).
- McGeoch, A. T., Trakselis, M. A., Laskey, R. A. & Bell, S. D. Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. *Nature Struct. Mol. Biol.* **12**, 756–762 (2005).
- Costa, A. *et al.* DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. *eLife* **3**, e03273 (2014).
- Graham, B. W., Schauer, G. D., Leuba, S. H. & Trakselis, M. A. Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding. *Nucleic Acids Res.* **39**, 6585–6595 (2011).
- Sun, J. *et al.* Structural and mechanistic insights into Mcm2–7 double-hexamer assembly and function. *Genes Dev.* **28**, 2291–2303 (2014).
- Samel, S. A. *et al.* A unique DNA entry gate serves for regulated loading of the eukaryotic replicative helicase MCM2–7 onto DNA. *Genes Dev.* **28**, 1653–1666 (2014).
- Sun, J. *et al.* Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2–7 bound to DNA. *Nature Struct. Mol. Biol.* **20**, 944–951 (2013).
- Costa, A. *et al.* The structural basis for MCM2–7 helicase activation by GINS and Cdc45. *Nature Struct. Mol. Biol.* **18**, 471–477 (2011).
- Hesketh, E. L. *et al.* DNA induces conformational changes in a recombinant human minichromosome maintenance complex. *J. Biol. Chem.* **290**, 7973–7979 (2015).
- Brewster, A. S. *et al.* Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. *Proc. Natl Acad. Sci. USA* **105**, 20191–20196 (2008).
- Bae, B. *et al.* Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog. *Structure* **17**, 211–222 (2009).
- Slaymaker, I. M. *et al.* Mini-chromosome maintenance complexes form a filament to remodel DNA structure and topology. *Nucleic Acids Res.* **41**, 3446–3456 (2013).
- Fletcher, R. J. *et al.* The structure and function of MCM from archaeal *M. thermoautotrophicum*. *Nature Struct. Mol. Biol.* **10**, 160–167 (2003).
- Froelich, C. A., Kang, S., Epling, L. B., Bell, S. P. & Enemark, E. J. A conserved MCM single-stranded DNA binding element is essential for replication initiation. *eLife* **3**, e01993 (2014).
- Fu, Y., Slaymaker, I. M., Wang, J., Wang, G. & Chen, X. S. The 1.8-Å crystal structure of the N-terminal domain of an archaeal MCM as a right-handed filament. *J. Mol. Biol.* **426**, 1512–1523 (2014).
- Liu, W., Pucci, B., Rossi, M., Pisani, F. M. & Ladenstein, R. Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain. *Nucleic Acids Res.* **36**, 3235–3243 (2008).
- Miller, J. M., Arachea, B. T., Epling, L. B. & Enemark, E. J. Analysis of the crystal structure of an active MCM hexamer. *eLife* **3**, e03433 (2014).
- Cuesta, I. *et al.* Conformational rearrangements of SV40 large T antigen during early replication events. *J. Mol. Biol.* **397**, 1276–1286 (2010).
- Vijayraghavan, S. & Schwacha, A. The eukaryotic Mcm2–7 replicative helicase. *Subcell. Biochem.* **62**, 113–134 (2012).
- Bochman, M. L., Bell, S. P. & Schwacha, A. Subunit organization of Mcm2–7 and the unequal role of active sites in ATP hydrolysis and viability. *Mol. Cell. Biol.* **28**, 5865–5873 (2008).
- Evrin, C. *et al.* The ORC/Cdc6/MCM2–7 complex facilitates MCM2–7 dimerization during prereplicative complex formation. *Nucleic Acids Res.* **42**, 2257–2269 (2014).
- Slaymaker, I. M. & Chen, X. S. MCM structure and mechanics: what we have learned from archaeal MCM. *Subcell. Biochem.* **62**, 89–111 (2012).
- Bochman, M. L. & Schwacha, A. The MCM complex: unwinding the mechanism of a replicative helicase. *Microbiol. Mol. Biol. Rev.* **73**, 652–683 (2009).
- Shima, N. *et al.* A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. *Nature Genet.* **39**, 93–98 (2007).
- Hardy, C. F., Dryga, O., Seematter, S., Pahl, P. M. & Sclafani, R. A. mcm5/cdc46-bob1 bypasses the requirement for the S phase activator Cdc7p. *Proc. Natl Acad. Sci. USA* **94**, 3151–3155 (1997).
- Bleichert, F., Botchan, M. R. & Berger, J. M. Crystal structure of the eukaryotic origin recognition complex. *Nature* **519**, 321–326 (2015).
- Enemark, E. J. & Joshua-Tor, L. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**, 270–275 (2006).
- Kang, S., Warner, M. D. & Bell, S. P. Multiple functions for Mcm2–7 ATPase motifs during replication initiation. *Mol. Cell* **55**, 655–665 (2014).
- Coster, G., Frigola, J., Beuron, F., Morris, E. P. & Diffley, J. F. Origin licensing requires ATP binding and hydrolysis by the MCM replicative helicase. *Mol. Cell* **55**, 666–677 (2014).
- Bell, S. D. & Botchan, M. R. The minichromosome maintenance replicative helicase. *Cold Spring Harb. Perspect. Biol.* **5**, a012807 (2013).
- Jenkinson, E. R. & Chong, J. P. Minichromosome maintenance helicase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert. *Proc. Natl Acad. Sci. USA* **103**, 7613–7618 (2006).

45. Gai, D., Zhao, R., Li, D., Finkelstein, C. V. & Chen, X. S. Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**, 47–60 (2004).
46. On, K. F. *et al.* Prereplicative complexes assembled *in vitro* support origin-dependent and independent DNA replication. *EMBO J.* **33**, 605–620 (2014).
47. Bruck, I. & Kaplan, D. L. The Dbf4-Cdc7 kinase promotes Mcm2–7 ring opening to allow for single-stranded DNA extrusion and helicase assembly. *J. Biol. Chem.* **290**, 1210–1221 (2015).
48. Bruck, I. & Kaplan, D. L. Cdc45 protein-single-stranded DNA interaction is important for stalling the helicase during replication stress. *J. Biol. Chem.* **288**, 7550–7563 (2013).
49. Fien, K. *et al.* Primer utilization by DNA polymerase α -primase is influenced by its interaction with Mcm10p. *J. Biol. Chem.* **279**, 16144–16153 (2004).
50. Eisenberg, S., Korza, G., Carson, J., Liachko, I. & Tye, B. K. Novel DNA binding properties of the Mcm10 protein from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **284**, 25412–25420 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank X. Li for providing programs in data collection, motion correction and framed-based analysis, and J. Wang for advices on modelling and model

refinement. We also thank the National Center for Protein Sciences (Beijing, China) for technical support with cryo-EM data collection and for computation resource. This work was supported by the Ministry of Science and Technology of China (2013CB910404 to N.G.), the National Natural Science Foundation of China (31422016 to N.G.), the Research Grants Council of Hong Kong (GRF664013 and HKUST12/CRF/13G to Yu.Z.) and the Hong Kong University of Science & Technology (B.-K.T.).

Author Contributions Yu.Z. purified sample; N.L. collected cryo-EM data (with J.L., Yi.Z. and W.L.), performed image processing, and analyzed structures. N.L., N.G. and M.Y. performed atomic modelling. N.L., Yu.Z., B.-K.T. and N.G. designed experiments, interpreted the structure and wrote the manuscript.

Author Information The cryo-EM density map has been deposited in the Electron Microscopy Data Bank (EMDB) under accession number EMD-6338; and the atomic model has been deposited in the Protein Data Bank (PDB) under accession number 3JA8. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.G. (ninggao@tsinghua.edu.cn), B.-K.T. (bt16@cornell.edu), or Yu.Z. (zhai@ust.hk).

METHODS

No statistical methods were used to predetermine sample size.

Yeast strain. One-step PCR-based approach⁵¹ with pTF272 (pFA6a-TEV-6×Gly-3×Flag-HphMX, Addgene) as DNA template was used to generate MCM4-TEV-3×Flag tagging modification in the W303-1a background strain. The resulting strain showed no growth defect compared to its parent W303-1a strain.

Sample purification. Forty litres of log-phase G1 yeast cells (3×10^7 – 4×10^7 cells per ml) were collected and processed for spheroplasting to isolate crude chromatin as described previously⁵² with the following modifications for a large-scale preparation. Spheroplasting was performed in 200 ml of spheroplasting buffer containing sufficient amount of lyticase that was purified from an *Escherichia coli* strain bearing lyticase expressing plasmid pUV5-GIS (gift from S. Gasser). The spheroplasts were lysed with extraction buffer EBX (50 mM HEPES/KOH, pH 7.5, 100 mM K-glutamate, 10 mM magnesium acetate, 0.25% Triton X-100, 3 mM ATP, 1 mM dithiothreitol (DTT), 1 mM EDTA, 2 mM NaF, 1 mM NaVO₄, 1 mM phenylmethanesulfonylfluoride (PMSF), 2 µg ml^{−1} pepstatin A and 1× protease inhibitor cocktail (Roche)). The lysate was layered onto the top of equal volume of EBX buffer containing 30% sucrose and centrifuged at 25,000g (Hitachi R20A2) for 15 min. To solubilize chromatin fractions, the crude chromatin was digested in 40 ml of freshly made benzonase buffer (50 mM HEPES/KOH, pH 7.5, 100 mM K-glutamate, 8 mM MgCl₂, 0.02% NP-40, 3 mM ATP, 1 mM EDTA, 2 mM NaF, 1 mM NaVO₄, 1 mM PMSF, 2 µg ml^{−1} pepstatin A and 1× protease inhibitor cocktail (Roche)) with 1 U µl^{−1} of benzonase (71206-3; Merck Biosciences) for 10 min at 37 °C, and then 1 h on ice. The suspension was then centrifuged for 20 min at 25,000g. The clear phase was recovered, and subjected to anti-Flag immunoprecipitation with 1 ml bed volume of washed anti-Flag M2 agarose (Sigma) at 4 °C for 2 h. Beads were recovered, and washed extensively with benzonase buffer and then tobacco etch virus (TEV) buffer (50 mM HEPES/KOH, pH 7.5, 100 mM K-glutamate, 8 mM MgCl₂, 0.02% NP-40, 3 mM ATP). MCM2–7 complexes were cleaved from the M2 agarose by incubation for overnight at 4 °C in TEV buffer with 100 U ml^{−1} of AcTEV protease (Life Technology). His-tagged TEV protease was removed by incubating the eluate with a TALON metal affinity resin (Clontech) for 30 min at 4 °C. The MCM2–7 complexes were then applied on the top of 20–40% glycerol gradient in buffer EBX with protease inhibitors. The glycerol gradient was centrifuged in a TLS-55 rotor (Beckman Optima TLX ultracentrifuge) at 175,000g for 6.5 h. The fractions were collected from the top of the gradient after centrifugation. The fractions containing the MCM2–7 double hexamers were pooled and processed for electron microscopy analysis.

Electron microscopy. The MCM2–7 double hexamer was concentrated by ultrafiltration to remove glycerol. Negative staining of the MCM2–7 double hexamer was performed with 2% uranyl acetate. Grids were examined using an FEI T12 microscope operated at 120 kV, and images were recorded using a 4k × 4k charge-coupled device (CCD) camera (UltraScan 4000, Gatan).

For cryo-grid preparation, 4 µl aliquots of samples were applied to a glow-discharged holey carbon grid (Quantifoil R2/2) coated with a thin layer of freshly prepared carbon, and cryo-freezing was performed with an FEI Vitrobot Mark IV (4 °C and 100% humidity). Grids were examined using an FEI Titan Krios operated at 300 kV, and images were recorded using a K2 Summit direct electron detector (Gatan) in counting mode, at a nominal magnification of 22,500×, which renders a final pixel size of 1.32 Å at object scale after post-magnification calibration, and with the defocus ranging from −1.5 to −2.5 µm. Images were collected under low-dose condition in a semi-automatic manner using UCSF-Image4 (written by X. Li and Y. Cheng). For each micrograph stack, a total of 32 frames were collected, with a dose rate of ~8.2 counts (~10.9 electrons) per physical pixel per second for an exposure time of 8 s.

Image processing. Initial 3D model from negatively stained particles was calculated using RELION⁵³ using a density cylinder as reference. For cryo-EM data, beam-induced motion correction at micrograph level was performed as previously described (written by X. Li)⁵⁴. Micrographs screening, automatic particle picking and normalization were done with SPIDER⁵⁵. Program of CTFFIND3 (ref. 56) was used to estimate the contrast transfer function parameters. The 2D, 3D classification and refinement were performed with RELION. A total of 347,801 particles (with a binning factor of two) from 2,230 micrographs were subjected to a cascade of 2D and 3D classification. Analysis of classification structures indicated that there is a C2-axis perpendicular to the cylinder axis of the MCM2–7 double hexamer, reflecting a symmetric arrangement of one single hexamer relatively to the other single hexamer by a simple 180° rotation. A final structurally homogeneous data set composed of 85,365 particles, as classification structures of them have reached to considerably higher resolution, in full window size (300 × 300) were used for high-resolution refinement with C2-symmetry imposed. From the orienta-

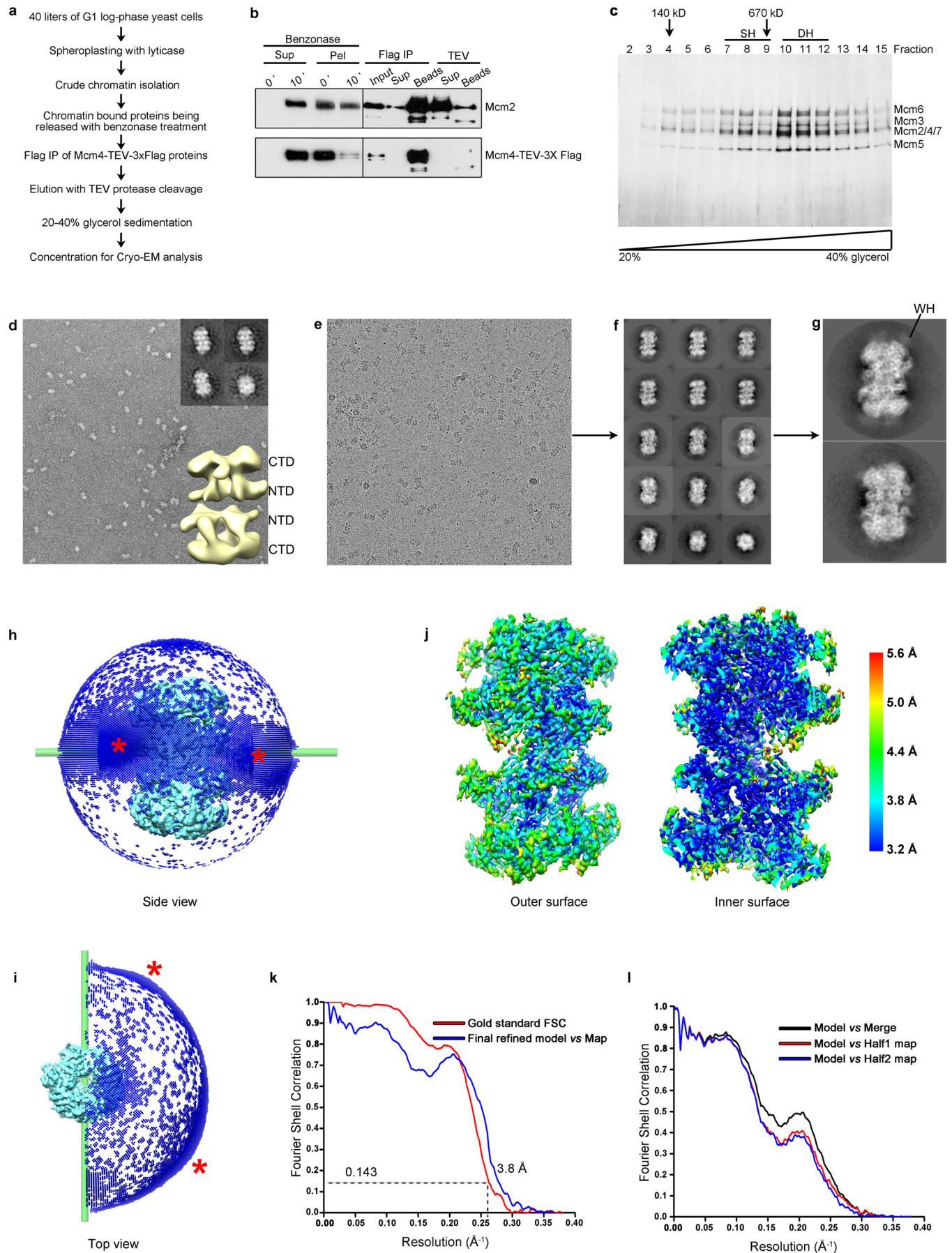
tion distribution (Extended Data Fig. 1h, i), there is a wide equator belt with a complete distribution of particles, along with two regions with relatively more particles. Nevertheless, this type of uneven distribution did not affect our final reconstruction, as particles from the equator belt have provided sufficient information for a complete sampling of the central slices in the Fourier space. Symmetry-free refinement was also performed, resulting in generally similar but slightly worse density maps. To improve the resolution further, different combinations of movie frames were used for motion correction and frame averaging. The first two frames had large motions, therefore, frames 3–16 were used to sum micrographs. To reduce interpolation errors, particles were rewindowed by offsetting translation parameters determined in the 3D refinement of last round, which improved the resolution to 4.6 Å (gold-standard FSC 0.143 criteria). The final round of refinement was performed with a soft-edged mask applied, resulting in a 4.3-Å map. After correction for the modulation transfer function of K2 detector, and map sharpening using post-processing options of RELION with a B-factor of −100 Å², the overall resolution of the final density map within the region defined by the soft mask is 3.8 Å for the overall map (Extended Data Fig. 1k), after correction of the effect of soft mask on the FSC curve⁵⁷. Local resolution map was estimated using *blocres* in Bsoft⁵⁸. From the local resolution map, peripheral regions are associated with worst resolution, while the core region is better than 3.5 Å. The statistics of the data collection and structural refinement is provided in Extended Data Table 1b.

Model building. Six monomers of the crystal structure of a chimaeric archaeal MCM (PDB code 4R7Y)³⁰ (*Sulfolobus solfataricus* NTD fused with *Pyrococcus furiosus* CTD) hexamer were manually docked to the density map of the MCM double hexamer using Chimera⁵⁹. The docking also confirmed the handedness of the density map. The rigid-body docking was performed by dividing the crystal structure of the monomer into four pieces (NTD-A, OB-fold, ZF and CTD) (Extended Data Fig. 4b, c). Sequence alignments of the yeast MCM proteins with crystal template were initially performed using BLAST⁶⁰ and manually adjusted according to the secondary structure prediction of these sequences (PSIPRED)⁶¹. The predicted secondary structural information of the eukaryotic subunit-specific sequences was used to assign the six MCM proteins into the cryo-EM density map. Initial atomic coordinates of the OB-fold subdomains and CTDs of MCM2–7 proteins were then generated using CHAINSAW⁶² in the CCP4 suite⁶³. Models were manually adjusted and built in Coot⁶⁴. Only minor changes were required for modelling the OB-fold subdomains and CTDs of the yeast MCM proteins owing to their high sequence identity to the template (Extended Data Fig. 2). The NTD-As of the yeast MCM proteins contain many sequence insertions, and the modelling of these sequences was similar to that described above, but involved multiple rounds of realignment of sequences and largely facilitated by the predicted secondary structure. In many cases, the modelling of NTD-A required complete retracing of the main-chain based solely on densities and secondary structural predictions. For regions independent of known template (eukaryotic-specific sequences, Extended Data Fig. 4), poly-alanine models were built first using Coot. Clearly resolved bulky residues (Phe, Tyr, Trp and Arg) were then used as markers to assign the primary sequences. As a result, we derived an atomic model of the MCM2–7 double hexamer, for ~80% of its sequences, from a near-atomic cryo-EM density map, integrated with structural information from other sources. Further model refinement was done by alternating rounds of model rebuilding in Coot and real-space refinement (phenix.real_space_refine)⁶⁵ in Phenix⁶⁶, with secondary structure and stereochemical constraints applied. Similar to a previous cryo-EM work with comparable 3.8-Å resolution⁵⁷, during the real-space refinement, knowledge-based restraints, including Ramachandran potentials and rotamer correction, were applied to ensure a proper balance between density-fitting and stereochemical and rotamer distributions. Owing to the resolution limitation, local densities at the ATP-binding sites could not unambiguously distinguish between ATP and ADP. For modelling purpose, ADP was docked to the active centres and similarly refined in Phenix. The atomic model was cross-validated according to previously described procedures^{68,69}. Specifically, the coordinates of the final model were randomly displaced by 0.2 Å using the PDB tools of Phenix. The displaced model was refined against the Half1 map (produced from a half set of all particles during refinement by RELION). The refined model from Half1 map was compared with the maps of Half1, Half2 in Fourier space to produce two FSC curves, FSC_{work} (model versus Half1 map) and FSC_{free} (model versus Half2 map), respectively (Extended Data Fig. 1l). Another FSC curve between the refined model from Half1 and the final density map (model versus merge) from all particles was also produced. As indicated by these curves, the agreement between FSC_{work} and FSC_{free} (no large separation) indicated that the model was not overfitted. MolProbity⁷⁰ (<http://molprobity.biochem.duke.edu/>) was used to evaluate the final model, and final statistics of the model was provided in Extended Data Table 1b. Notably, applica-

tion of knowledge-based restraints during the real-space refinement has improved the stereochemical and rotamer statistics of the model. Comparisons of the representative density with the atomic model for selected areas are shown in Extended Data Fig. 6 and Supplementary Video 2.

Gravity centres of individual domains (Fig. 1) were determined with segmented maps of the conserved core regions of these domains (minus variable loops and linkers) using SPIDER. To determine the cylinder axis of the hexamer, a plane perpendicular to the axis was determined by least-square fitting of six centres of the OBs in Chimera. Pymol⁷¹ and Chimera were used for structural analysis and figure preparation. Interface areas of the intersubunit and inter-hexamer interactions were calculated by PISA⁷², and provided in Extended Data Table 1a.

51. Janke, C. *et al.* A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947–962 (2004).
52. Zhai, Y., Yung, P. Y., Huo, L. & Liang, C. Cdc14p resets the competency of replication licensing by dephosphorylating multiple initiation proteins during mitotic exit in budding yeast. *J. Cell Sci.* **123**, 3933–3943 (2010).
53. Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
54. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
55. Shaikh, T. R. *et al.* SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nature Protocols* **3**, 1941–1974 (2008).
56. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
57. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
58. Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).
59. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
60. Mount, D. W. Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc.* **2007**, pdb.top17 (2007).
61. Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **41**, W349–W357 (2013).
62. Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643 (2008).
63. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
64. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
65. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
66. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
67. Wang, Z. *et al.* An atomic model of brome mosaic virus using direct electron detection and real-space optimization. *Nature Commun.* **5**, 4808 (2014).
68. Zhao, M. *et al.* Mechanistic insights into the recycling machine of the SNARE complex. *Nature* **518**, 61–67 (2015).
69. Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).
70. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
71. Schrodinger, L. L. C. *The PyMOL Molecular Graphics System v.1.3r1* (2010).
72. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
73. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98 (1999).
74. Wei, Z. *et al.* Characterization and structure determination of the Cdt1 binding domain of human minichromosome maintenance (Mcm) 6. *J. Biol. Chem.* **285**, 12469–12473 (2010).

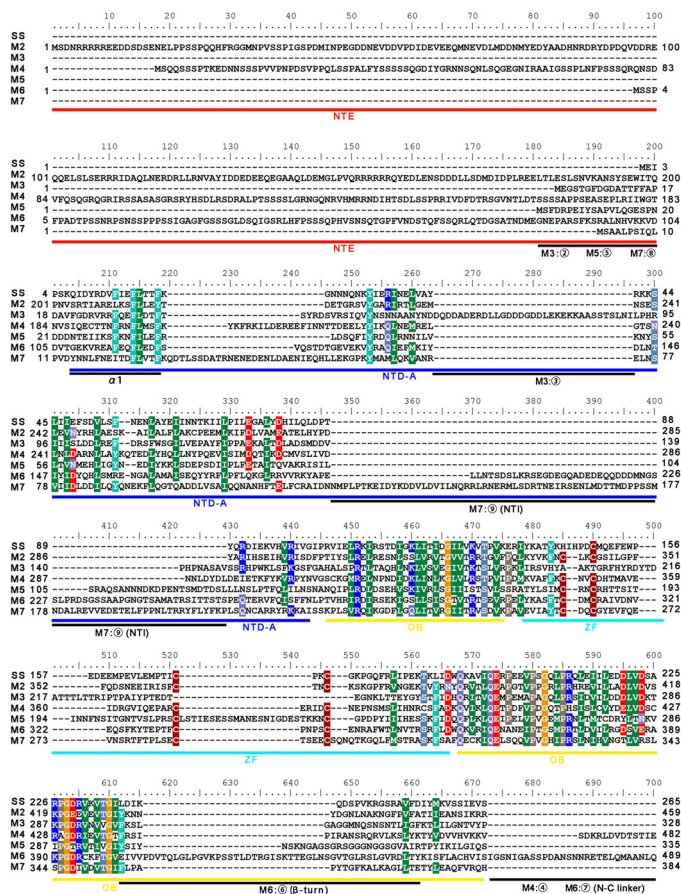


Extended Data Figure 1 | MCM2–7 double hexamer purification and structural determination.

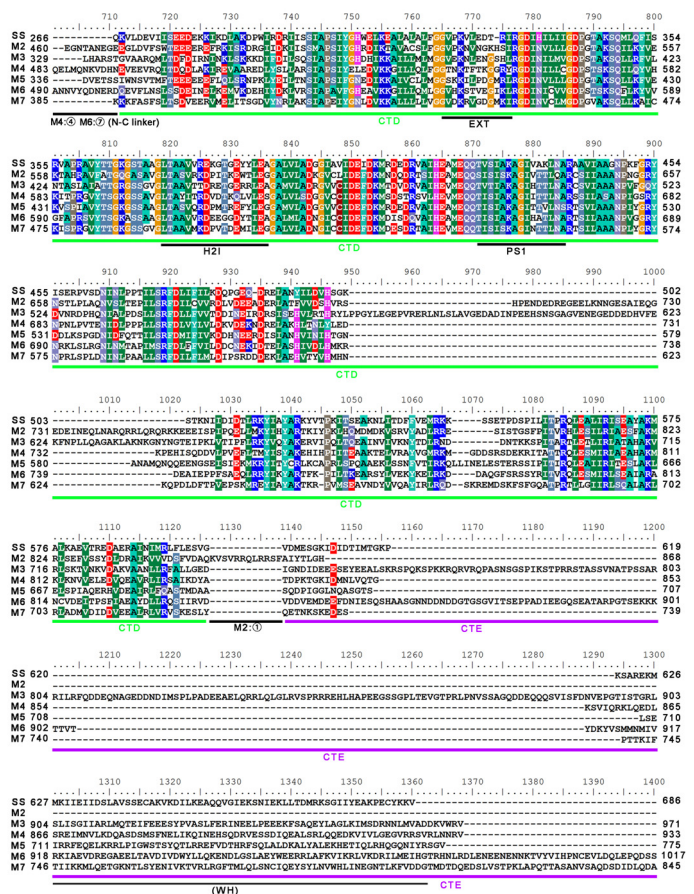
a, A flowchart of the procedure for MCM2–7 double hexamer purification from G1 chromatin of the yeast strain MCM4-TEV-3×Flag. **b**, Fractions taken were analysed by SDS–PAGE and immunoblotting of the indicated MCM subunits. **c**, The eluted MCM2–7 complexes were subjected to 20–40% glycerol gradient sedimentation at 175,000g for 6.5 h. Collected fractions were analysed by SDS–PAGE and visualized by silver staining. Molecular size markers used are: ALP 140 kDa and thyroglobulin 670 kDa. Fractions 10–12 were pooled and concentrated for cryo-EM analysis. **d**, A representative raw micrograph of the negatively stained MCM2–7 double hexamer. Representative 2D class averages of negatively stained particles produced by reference-free classification are shown at the top-right corner. The initial 3D model generated using RELION is shown at the bottom-right corner. **e**, A representative raw micrograph of cryo-EM data. **f**, Representative 2D class averages of cryo-EM particles from reference-free classification. **g**, Two typical side views of the average images from **f**, in enlarged

forms, highlighting well-resolved secondary structure elements. Extra densities with poor quality on the two ends of double hexamer could be attributed to the flexible winged-helix motifs (WH) within the CTEs of MCM proteins.

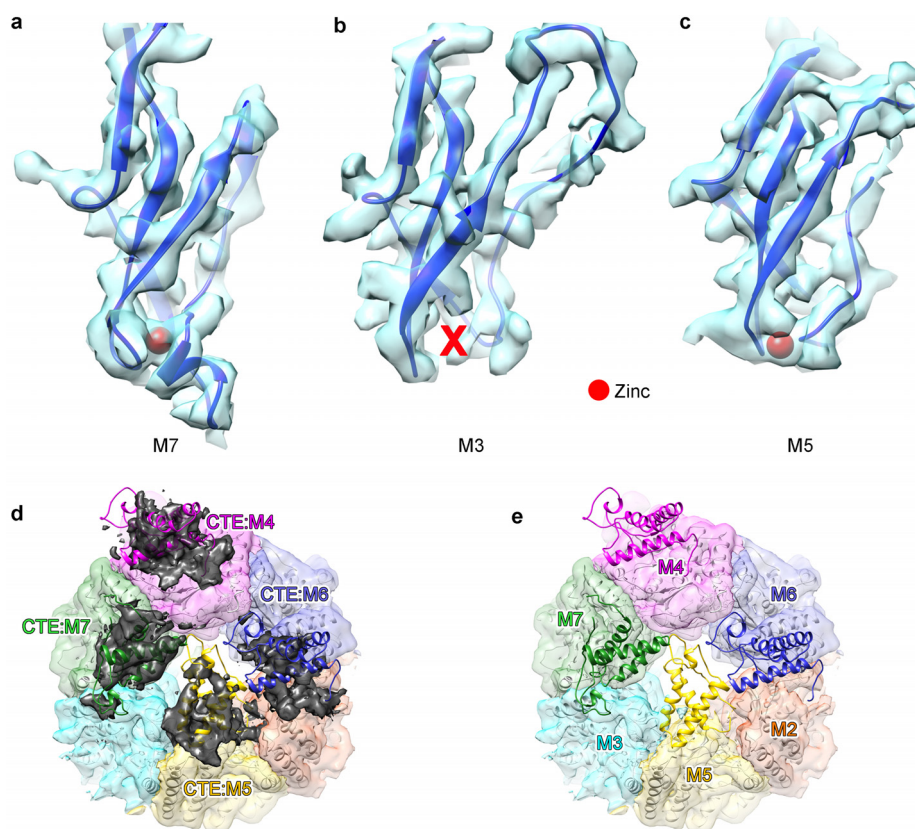
h, **i**, Distribution of particle orientations in the last round of structural refinement, showing in side (**h**) and top (**i**) views. The heights of blue cylinders at different projection directions on the surface of a hemisphere are proportional to their particle numbers. Two areas (red asterisks) of a dense equator belt are slightly enriched with particles. **j**, The density map of the MCM2–7 double hexamer (sharpened) is shown in two views, for the outer (left) and inner (right) surfaces. The map is colour-coded to indicate the range of the local resolution. **k**, Fourier shell correlation (FSC) curves for the final 3D density map after RELION-based post-processing (red, gold-standard FSC), and for the cross-examination between final atomic model and the 3D density map (blue, final refined model versus map). At a FSC 0.143 cut-off, the overall resolution for the map is 3.8 Å. **l**, FSC curves for the atomic model cross-validation. See Methods for details.



Extended Data Figure 2 | Sequence alignment of the MCM2–7 proteins. The sequences of archaeal (SS, *Sulfolobus solfataricus*) and yeast (*Saccharomyces cerevisiae*) MCM proteins were aligned using BioEdit⁷³. The alignment was further adjusted manually according to the secondary structure prediction and 3D structural alignment. Conserved hairpins and loops are

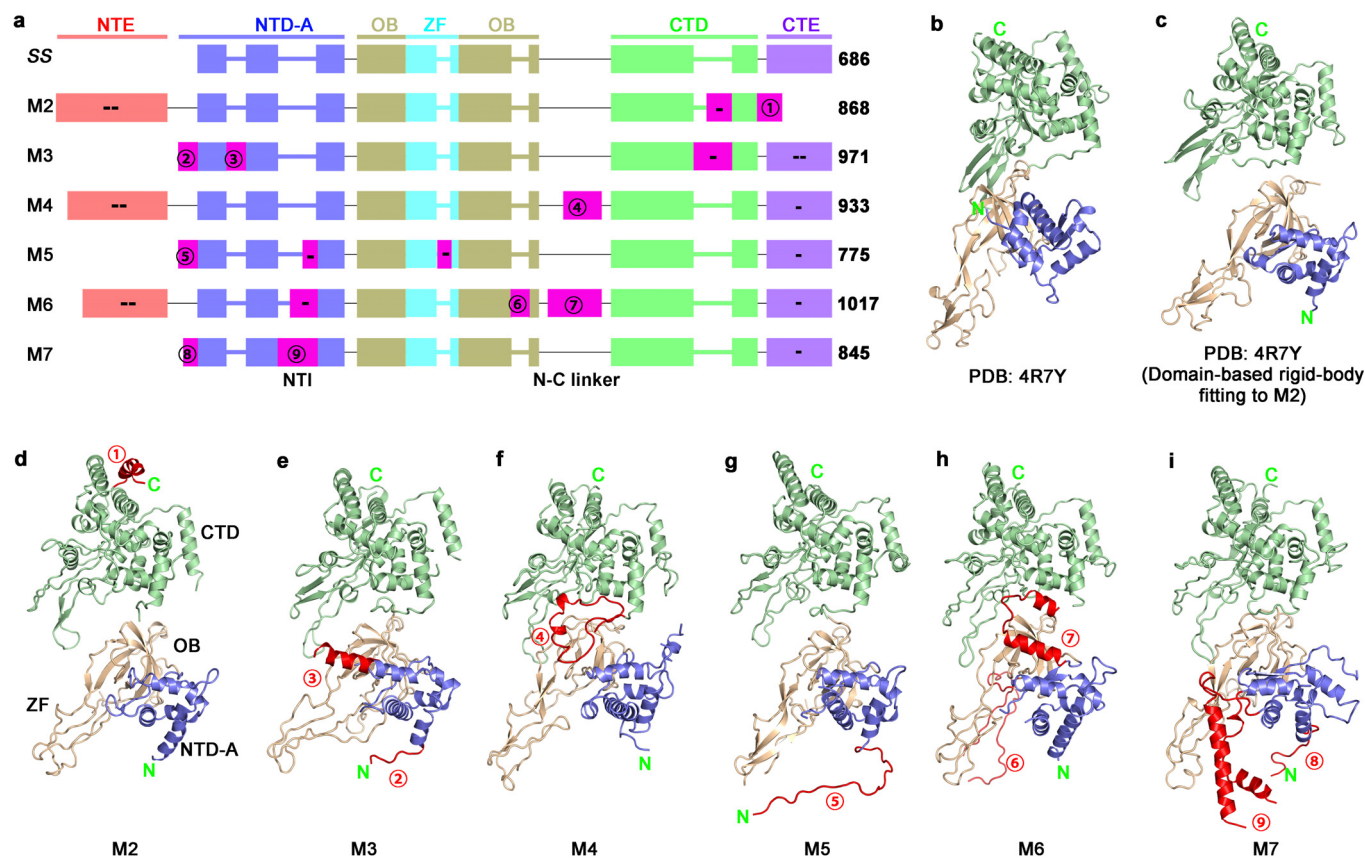


labelled (H2I, EXT, PS1 and β -turn). CTEs were aligned by the predicted secondary elements in the winged-helix (WH) motifs. Eukaryote-specific sequences (numbered 1–9) well resolved in our structure as in Extended Data Fig. 4a are labelled.



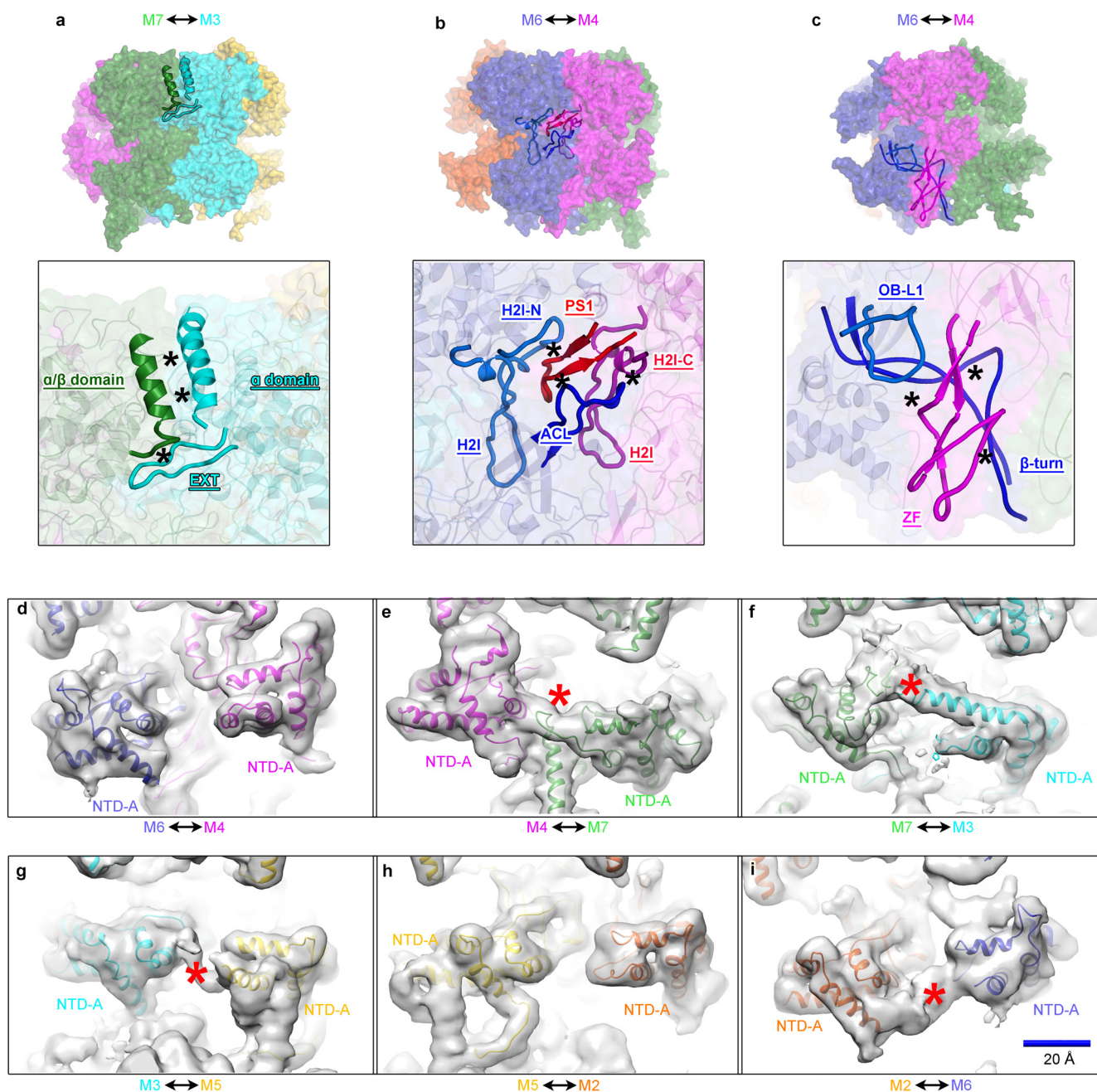
Extended Data Figure 3 | Structural diversity of ZFs and structural flexibility of CTEs in the MCM2-7 double hexamer. a–c, Ribbon representation of ZF motifs of MCM7 (a), MCM3 (b) and MCM5 (c), superimposed with sharpened density map (transparent cyan) at 4σ contour level. Positions of zinc are denoted by red balls. Zinc-binding was not observed in the ZF of MCM3. d, e, Surface representation of the density map (unsharpened, at 1σ

contour level) superimposed with colour-coded atomic structures for each MCM subunit, viewed from the CTD ring. Four segmented extra densities are coloured in deep grey (d), with tentative fitting of a winged-helix motif from a crystal structure (PDB code 2KLQ)⁷⁴ into these four density pieces. e, Same as d, but displayed without extra densities.



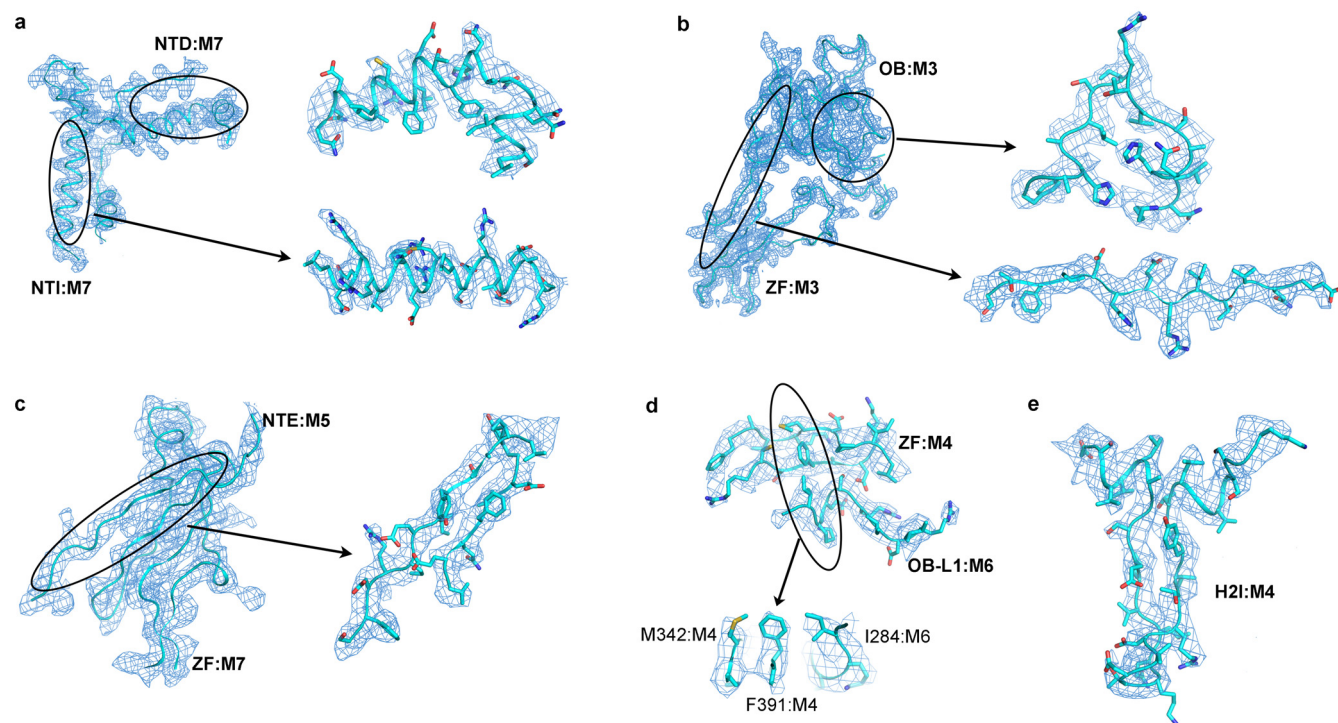
Extended Data Figure 4 | Subunit-specific structural features of the MCM2-7 subunits. **a**, Schematic illustration of domain organization and subunit-specific features of MCM2-7 subunits, with comparison to the archaeal MCM (SS, *Sulfolobus solfataricus*) (see also Extended Data Fig. 2). Numbered regions correspond to numbered extensions and insertions highlighted in **d-i**. '-' symbols denote corresponding regions with reliable densities to trace the main chain direction, but not sufficient for atomic modelling. '--' symbols denote sequences with highly disordered densities.

b, c, A protomer of the crystal structure of a chimaeric archaeal MCM hexamer structure (PDB code 4R7Y)³⁰ used as the template for modelling. The archaeal MCM was aligned globally (**b**) or domain-based flexibly fitted (**c**) to the atomic model of MCM2. **d-i**, Side-by-side structural comparison of MCM2-7 proteins, with MCM3-7 globally aligned to the atomic model of MCM2. The well-resolved insertions and extensions of each MCM subunit (**d-i**) are numbered and coloured in red.



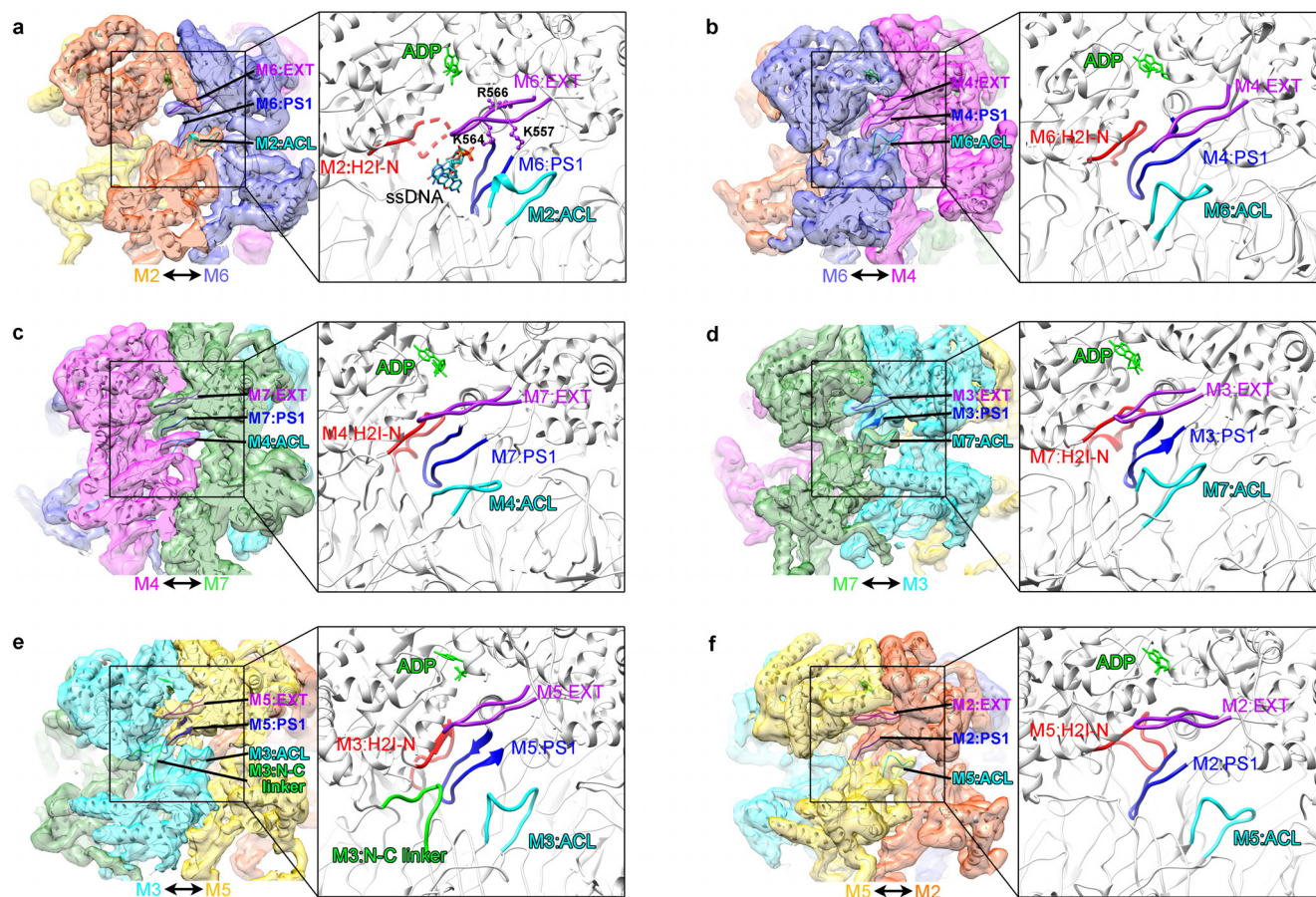
Extended Data Figure 5 | Intersubunit interactions in the MCM2-7 single hexamer. **a**, Interactions at the CTD ring exemplified by the 7:3 interface. The EXT hairpin of MCM3 facilitates the packing of one helix (the α -linker of the α/β subdomain) from MCM7 with another helix (located at the α subdomain of the CTD) from MCM3. **b**, Interactions at the neck region, as exemplified by the 6:4 interface. PS1-HP of MCM4 is sandwiched between ACL and H2I-N (N-terminal loop/helix of H2I) of MCM6. At the same time, ACL of MCM6 also interacts with H2I-C (C-terminal helix of H2I) of MCM4. **c**, Interactions at the NTD ring exemplified by the 6:4 interface. The first loop of OB (OB-L1) that flanks NTD-A, and the extended β -turn loop from MCM6

form a cradle for docking the ZF from MCM4. Asterisks mark sites of strong interactions. **d–i**, Zoomed-in views of intersubunit interactions between NTD-As of each adjacent MCM pair. The unsharpened density map (transparent grey), contoured at the 2.7σ level, is superimposed with the atomic model. Four of the six MCM proteins (3, 5, 6 and 7) contain NTIs at varying locations of their NTD-As (see also Extended Data Fig. 4). Only the NTI of MCM7 is modelled in our structure. Superimposition of the atomic model with the density map indicates that these NTIs all interact with the NTD-As of the adjacent subunit on the left. Extra densities indicating interactions are marked by red asterisks.



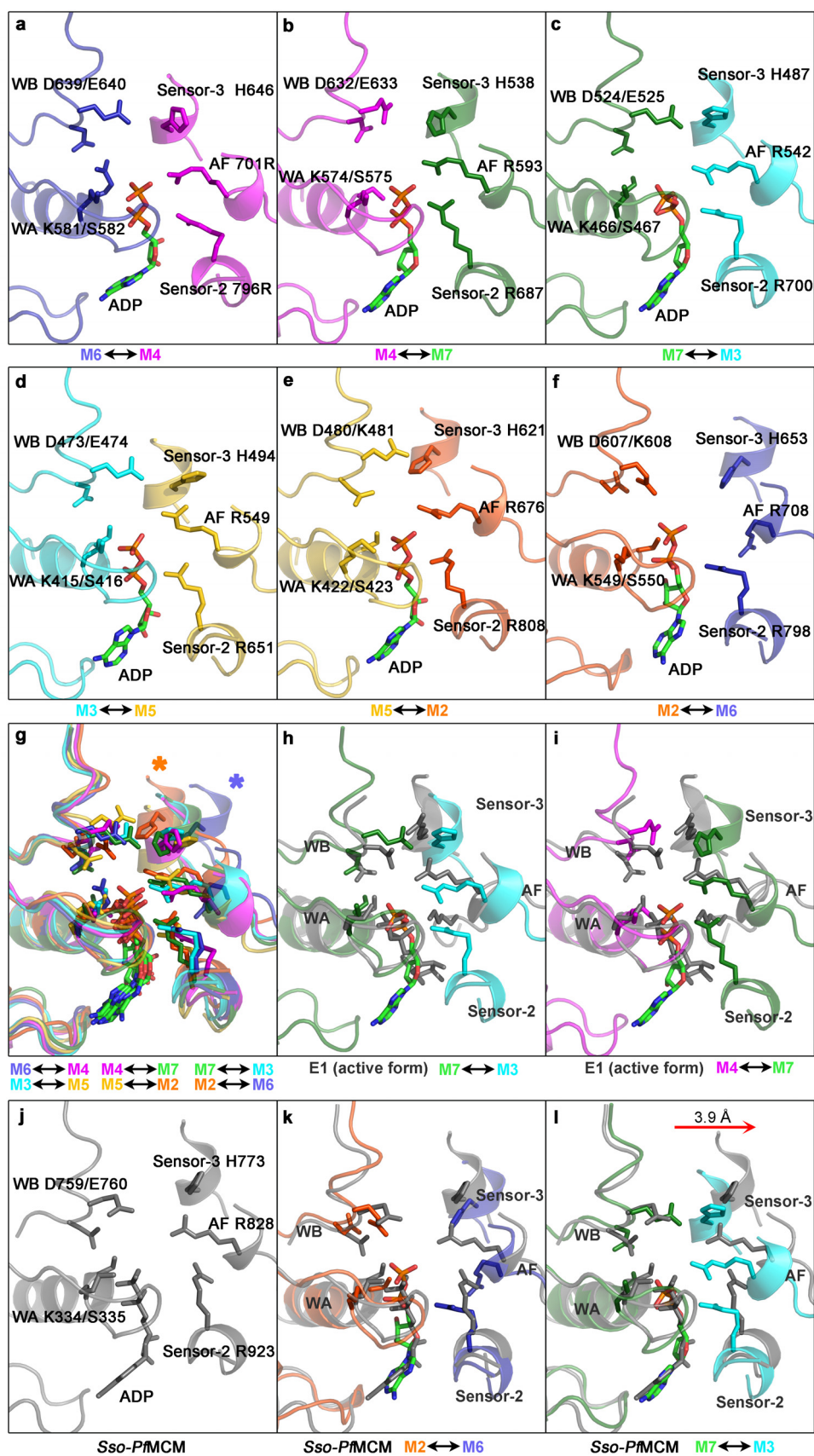
Extended Data Figure 6 | Cryo-EM densities for different regions of the MCM2-7 double hexamer. **a**, Electron microscopy density map (cyan mesh) superimposed with atomic model for NTD-A of MCM7. Two representative α -helices with side chains (right) were displayed in stick representation. **b**, Electron microscopy density map for the OB and ZF of MCM3. A representative loop of the OB and a strand connecting the OB and ZF with side chains are shown on the right. **c**, A representative region of inter-hexamer

interaction, highlighting the interactions between the β -strands of MCM5-NTE and MCM7-ZF. **d**, A representative region of intersubunit interaction (MCM4-MCM6), highlighting the hydrophobic interaction between Met342, Phe391 of MCM4 and neighbouring Ile284 of MCM6. **e**, A representative region of conserved hairpin loops, highlighting H2I of MCM4. Segmented density maps in all panels are displayed at the $5-6\sigma$ contour level.



Extended Data Figure 7 | A unique side channel between MCM2 and MCM6. **a–f**, Outer surface representation of the six subunit interfaces within MCM2–7 single hexamer. **a**, A unique side channel in the neck region of the M2–M6 interface. The boxed region is shown in a zoomed-in view (right) with individual components (H2I-N, EXT, PS1 and ACL) coloured individually. The size of this side channel is large enough to act as a pore for ssDNA exiting

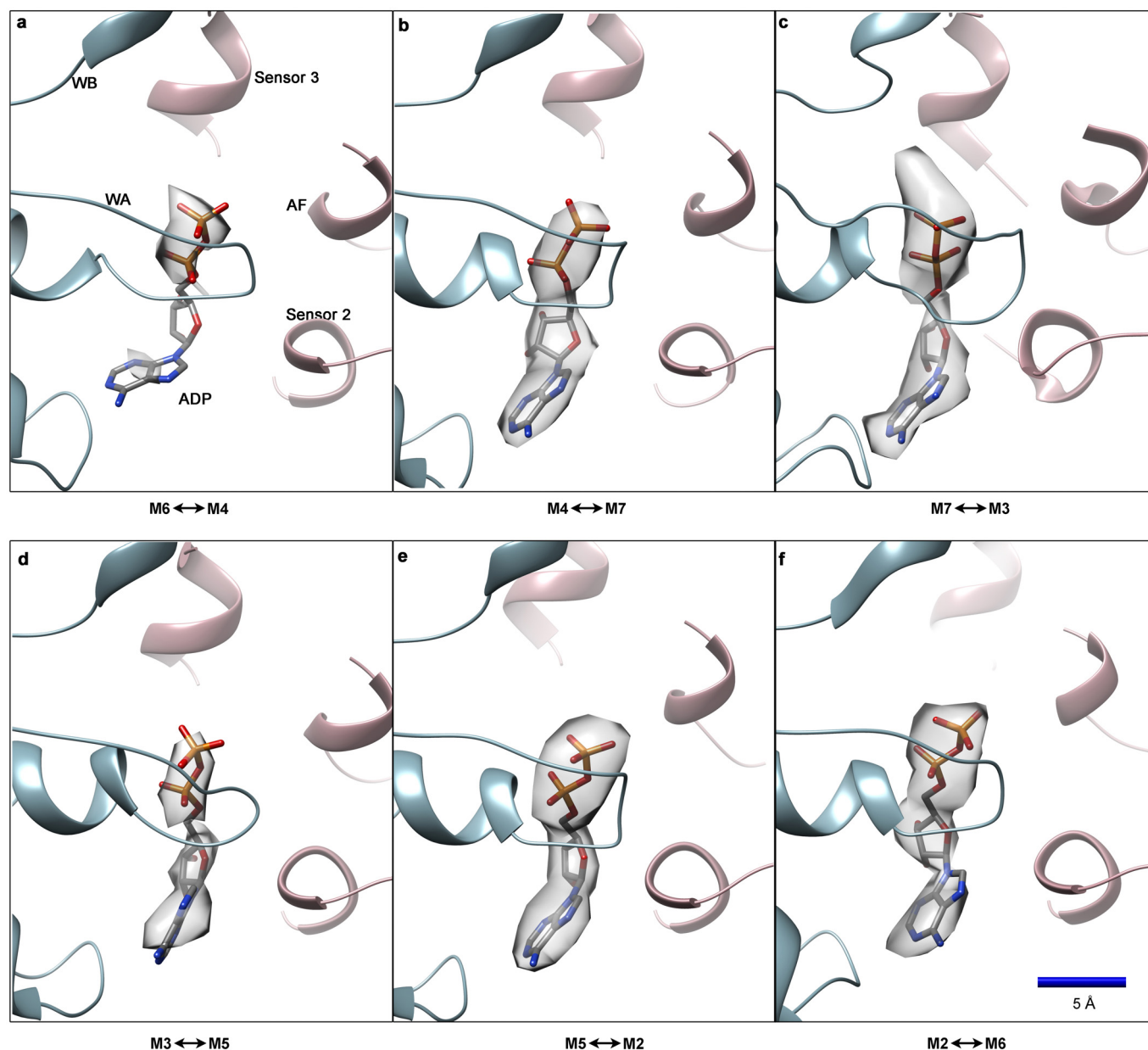
from the central channel during DNA unwinding along with basic residues (Arg566, Lys557 and Lys564) of the EXT hairpin from MCM6. The H2I-N is partially disordered. **b–f**, Same as **a**, but at different subunit interfaces. In the case of the 3:5 interface (**e**), the N–C linker of MCM3 also contributes to the blocking of the channel.



Extended Data Figure 8 | ATP-binding site configuration at MCM

intersubunit interfaces. **a–f**, Zoomed-in views of ATP-binding sites for each MCM dimer. The Walker A and B (WA and WB, respectively) residues of the left subunit, and sensor 3, sensor 2 and arginine finger (AF) residues of the right subunit, are shown in stick model. **g**, Superimposition of all six active centres. The sensor 3 residues of MCM2 (orange asterisk) and MCM6 (blue asterisk) in the 5:2 and 2:6 dimers display sharply different configurations,

resulting in two relatively loose centres. **h, i**, Superimposition of two representative compact ATPase centres (dimers of 7:3 and 4:7) with that of E1 hexameric helicase (active form)⁴⁰. **j**, The ATPase centre (inactive conformation) of an archaeal MCM (PDB code 4R7Y)³⁰. **k, l**, Superimposition of **j** with the centres of 2:6 (**k**) and 7:3 (**l**). Walker A and B motifs are used as a reference for alignment in all panels. **l**, A large shift in the sensor 3 of MCM3 is shown by red arrow, compared with the inactive conformation.



Extended Data Figure 9 | Nucleotide occupancy at the six ATPase centres of MCM2-7 single hexamer. a–f, Zoomed-in views of the active centres for all MCM subunit pairs. The conserved ATPase elements of the active centres are labelled. Segmented nucleotide densities at a contour level of 5.5σ were

superimposed (transparent grey). Note that nucleotide occupancies at the centres of 6:4 and 3:5 are relatively low. For the 7:3 dimer, there seems to be extra density for γ -phosphate or Mg^{2+} , but could not be confirmed at the current resolution (3.8 Å). Nucleotides were modelled using ADP.

Extended Data Table 1 | Statistics of structural determination, model refinement and interface analysis.

a

Buried surface areas for Inter-subunit interactions							
Subunits	Buried surface Total (Å ²)	Ratio and number of residues involved		Buried surface (Å ²)			
		Subunit 1	Subunit 2	Tier 1	Tier 2	Tier 3	Other (N/N, N/ZF, ZF/ZF)
M6/M4	4028	11.5%/114	11.1%/121	1011	1129	1800	230
M4/M7	4122	11.7%/111	10.4%/117	1611	994	1045	591
M7/M3	3825	9.8%/107	11.1%/123	1434	899	893	702
M3/M5	3884	11.4%/107	10.7%/113	1610	1198	994	271
M5/M2	3587	10.1%/103	10.3%/108	1397	1282	999	59
M2/M6	2886	8.6%/78	7.9%/96	1104	869	943	8

Buried surface areas for Inter-hexamer interactions (Å ²)										
	Total	5:7'/7:5'	3:5'/5:3'	2:6'/6:2'	3:7'/7:3'	3:3'	6:6'	5:5'	4:5'/5:4'	2:4'/4:2'
Total	6387	3284	979	847	625	270	114	102	135	30
ZF:ZF'	981		442	293			114	102		30
N:N'	2272	1487			538	111			136	
N:ZF'/ZF:N'	2531	1748	537		87	159				
β-turn/β-turn'	764	210		554						

Buried surface at the hexamer interface contributed by each subunit						
Subunit	M2	M3	M4	M5	M6	M7
Buried surface (Å ²)	877	1874	166	4500	961	3909

b

Data collection	
Electron microscope	Titan Krios
Voltage (kV)	300
Electron detector	K2 camera
Electron dose (e ⁻ /Å ²)	50 (32 frames)/22 (frame 3-16)
Pixel size (Å)	1.32
3D Reconstruction	
Particles for final refinement	85,366
Resolution of unmasked map (Å)	4.3
Resolution of masked map (Å)	3.8
Map sharpening B-factor (Å ²)	-100
Model composition	
Peptide chains	12
Residues	7,572
Ligands (ADP)	12
R.m.s. deviations	
Bonds length (Å)	0.0091
Bonds angles (°)	1.36
Ramachandran plot	
Favored (%)	91.0
Outliers (%)	1.1
Validation	
Molprobit score	2.45
Rotamer outliers (%)	0.06

a. Calculated surface areas of intersubunit and inter-hexamer interfaces. The calculation was done using PISA⁷². At the hexamer interface, there are 25 and 10 residues not built (owing to the structural disorder) for the ZF of MCM5 and the β-turn loop of MCM6, respectively. Therefore, the actual contribution of the MCM5-ZF and MCM6-β-turn to the interhexamer interaction could be much larger. **b.** Statistics of data processing and model refinement.

A giant protogalactic disk linked to the cosmic web

D. Christopher Martin¹, Mateusz Matuszewski¹, Patrick Morrissey¹, James D. Neill¹, Anna Moore², Sebastiano Cantalupo³, J. Xavier Prochaska^{4,5} & Daphne Chang[‡]

The specifics of how galaxies form from, and are fuelled by, gas from the intergalactic medium remain uncertain. Hydrodynamic simulations suggest that ‘cold accretion flows’—relatively cool (temperatures of the order of 10^4 kelvin), unshocked gas streaming along filaments of the cosmic web into dark-matter halos^{1–3}—are important. These flows are thought to deposit gas and angular momentum into the circumgalactic medium, creating disk- or ring-like structures that eventually coalesce into galaxies that form at filamentary intersections^{4,5}. Recently, a large and luminous filament, consistent with such a cold accretion flow, was discovered near the quasi-stellar object QSO UM287 at redshift 2.279 using narrow-band imaging⁶. Unfortunately, imaging is not sufficient to constrain the physical characteristics of the filament, to determine its kinematics, to explain how it is linked to nearby sources, or to account for its unusual brightness, more than a factor of ten above what is expected for a filament. Here we report a two-dimensional spectroscopic investigation of the emitting structure. We find that

the brightest emission region is an extended rotating hydrogen disk with a velocity profile that is characteristic of gas in a dark-matter halo with a mass of 10^{13} solar masses. This giant protogalactic disk appears to be connected to a quiescent filament that may extend beyond the virial radius of the halo. The geometry is strongly suggestive of a cold accretion flow.

We observed the UM287 filament with the Palomar Cosmic Web Imager (PCWI)⁷, an integral field spectrograph that is designed for low surface brightness measurements using a $40'' \times 60''$ reflective image slicer with twenty-four $40'' \times 2.5''$ slices. For these observations, the spectrograph covered a range of 3,940–4,110 Å with slit-width-limited resolution $\Delta\lambda \approx 2.5$ Å around the redshifted Lyman- α (Ly α) line. The methodology and details of our observations and data analysis are discussed in the Methods and described extensively elsewhere^{8,9}. Data reduction of the PCWI observations resulted in an adaptively smoothed data cube (consisting of the co-added sum of all the exposures) of dimensions right ascension, declination, and wavelength

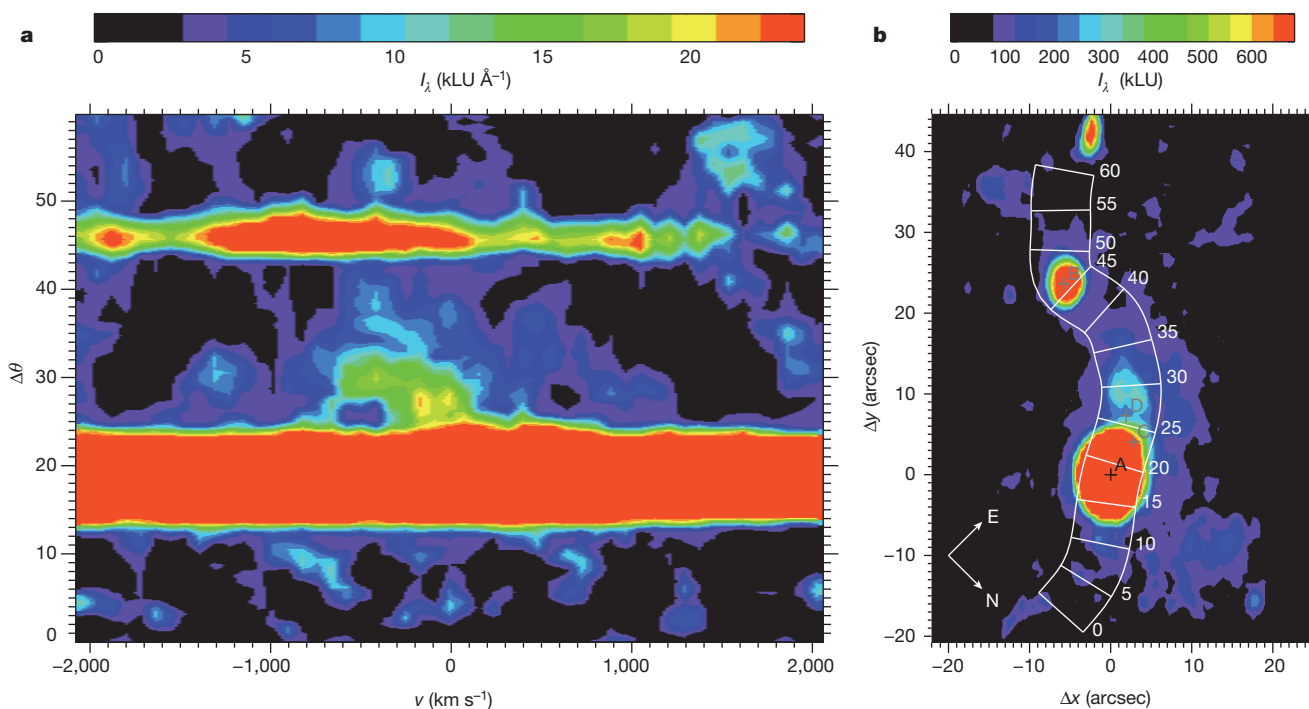


Figure 1 | Spectral image and pseudo-slit spectrum of the QSO UM287 field. **a**, Spectral image of the pseudo-slit (gridded region) in **b**, displayed in velocity with respect to the systemic velocity and arcseconds ($\Delta\theta$) with respect to the reference (QSO source A, which is QSO UM287) position ($z = 2.279$). Flux density I_λ is measured in kilo line units per Å, where $1 \text{ kLU } \text{\AA}^{-1} = 1,000 \text{ photon cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ or $1.18 \times 10^{-19} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ arcsec}^{-2}$. The signal-to-

noise ratio exceeds 7.0. Sources near QSO A are plotted. **b**, Narrow-band image generated from the PCWI data cube by summing the flux density I_λ over the band 3,970–4,000 Å showing the QSO UM287 (source A), the nearby, fainter QSO (source B), and the bright filament of Ly α emission extending from UM287. Emission sources C and D are also shown. The data cube is summed over 7.5 arcsec perpendicular to each position along the pseudo-slit to form the image in **a**.

¹Cahill Center for Astrophysics, California Institute of Technology, 1216 East California Boulevard, Mail code 278-17, Pasadena, California 91125, USA. ²Caltech Optical Observatories, Cahill Center for Astrophysics, California Institute of Technology, 1216 East California Boulevard, Mail code 11-17, Pasadena, California 91125, USA. ³ETH Zurich, Institute for Astronomy, Wolfgang-Pauli-Strasse 27 8093, Zurich, Switzerland. ⁴Department of Astronomy and Astrophysics, University of California, 1156 High Street, Santa Cruz, California 95064, USA. ⁵University of California Observatories, Lick Observatory, 1156 High Street, Santa Cruz, California 95064, USA. [‡]Deceased.

(RA, dec., λ). The bright extended filament is clearly detected in the data (Fig. 1): it is fainter far from QSO UM287 and shows a relatively narrow line (velocity) width (Fig. 1). Channel cuts through the smoothed data cube (Extended Data Figs 2, 4, 5) illustrate a disk morphology, and demonstrate that neither the QSO subtraction nor the presence of a nearby line and continuum object (source C) are responsible for the disk emission that is closest to QSO UM287 (see Methods). To reveal the spatial structure of the emission, we generated a new narrow-band window that follows the velocity-shear structure (Fig. 2).

The narrow-band and spectral images reveal an extended gaseous disk that is offset from, and illuminated by, the nearby quasi-stellar object QSO UM287 (which we call source A). The disk has a diameter of approximately 125 proper kiloparsecs (pkpc), with a central deficit with a diameter of about 25 pkpc. It is inclined by approximately 70° to the plane of the sky. The 2D velocity profile is well fitted by a rotating disk in a Navarro–Frenk–White (NFW) dark-matter halo with mass $\log_{10} M_h = 13.1^{+0.6}_{-0.3}$, where M_h is the mass of the halo in solar masses (M_\odot), halo concentration $c = 3^{+5}_{-2}$ (unitless), circular velocity (the maximum rotation velocity of the disk, at the virial radius) $v_c \approx 500 \text{ km s}^{-1}$, and virial radius $R_{\text{vir}} = 225 \text{ pkpc}$. Using a pseudo-slit (Fig. 2), we extracted a 1D velocity profile (Fig. 3), which is also well fitted by the same model, with consistent inferred mass

$\log_{10} M_h = 13.1^{+0.5}_{-0.3}$ and halo concentration $c = 5^{+7}_{-3}$. Both fits include the effects of slit averaging, and the instrument and seeing point spread functions (PSFs) that are measured using QSO A. The velocity profile flattens at the lower end (in detector coordinates), but continues to move blueward slowly at the upper end beyond the disk. The upper end of the disk smoothly transitions into a filament of emission that has a moderate velocity shear of about $100\text{--}150 \text{ km s}^{-1}$ over a length of at least 125 pkpc. An extended disk of gas provides a natural explanation for the unusual brightness of the QSO UM287 nebula, which when modelled as a cosmic web filament⁶ of typical column density required very high clumping factors. The clear disk-like morphology and kinematics of the nebula provide evidence that the Ly α spectral line is probing *in situ* kinematics^{10,11}.

We model the illumination of the disk by QSO source A and use simple arguments to constrain the geometry (Fig. 4). We assume that this QSO emits symmetrically into two coaxial cones, each with a solid angle of $\pi \text{ sr}$, with the rear cone illuminating the disk, and the front cone including our line of sight. We estimate that under these conditions the QSO can irradiate the disk and filament if the disk is more than 65 pkpc behind the QSO.

We used the ‘Cloudy’ nebular emission code¹² to model the disk and constrain the gas column density. We find that the gas is optically thin

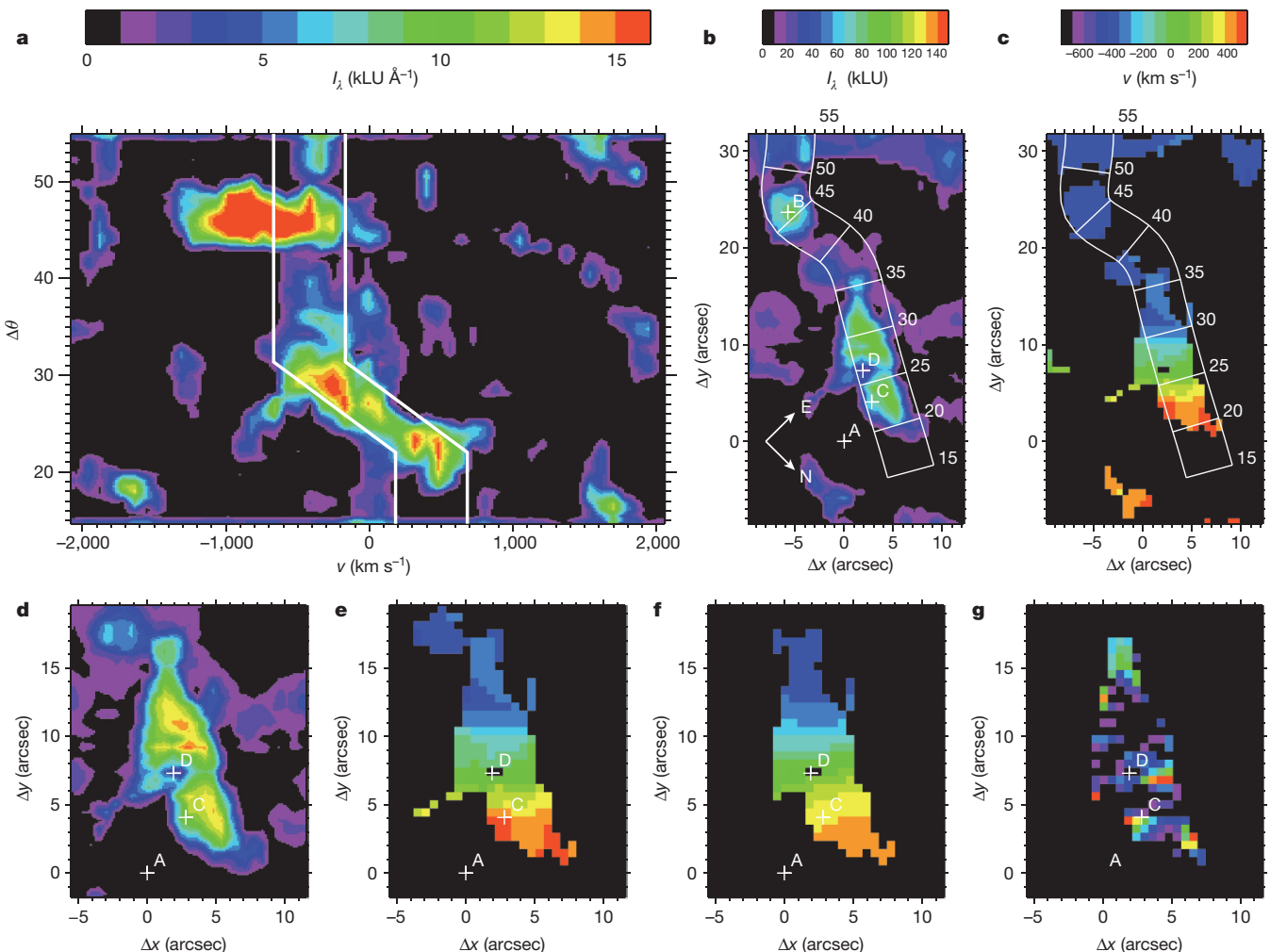


Figure 2 | Spectral image and narrow-band image created from a sheared-velocity window. **a**, Spectral image created with pseudo-slit (gridded region in **b**) and full-field subtraction. v_0 is the systemic velocity of the system. **b**, Narrow-band image formed using sheared, $\Delta v = 500 \text{ km s}^{-1}$ window (bold white lines in **a**). The image and spectrum are consistent with a large (15 arcsec or 125 pkpc), tilted (position angle of 15°), inclined (70°) disk with $v_c \approx 450 \text{ km s}^{-1}$ and a central deficit (at grid position (station) 26 arcsec, that is,

the grid location along slit). The blueshifted end of the disk transitions to an extended filament. **c**, Mean velocity in the sheared-velocity (kinematic) window (trimmed to show velocities only where emission exceeds 20% of the maximum value). **d**, **e**, Zooms of **b** and **c**, respectively. **f**, 2D velocity model that is compared with **c**. Panel **d** has the same colour scale as **b** and panels **e** and **f** have the same colour scale as **c**. **g**, χ^2 map of the 2D model fit residuals; full-scale $\chi^2 = 5$.

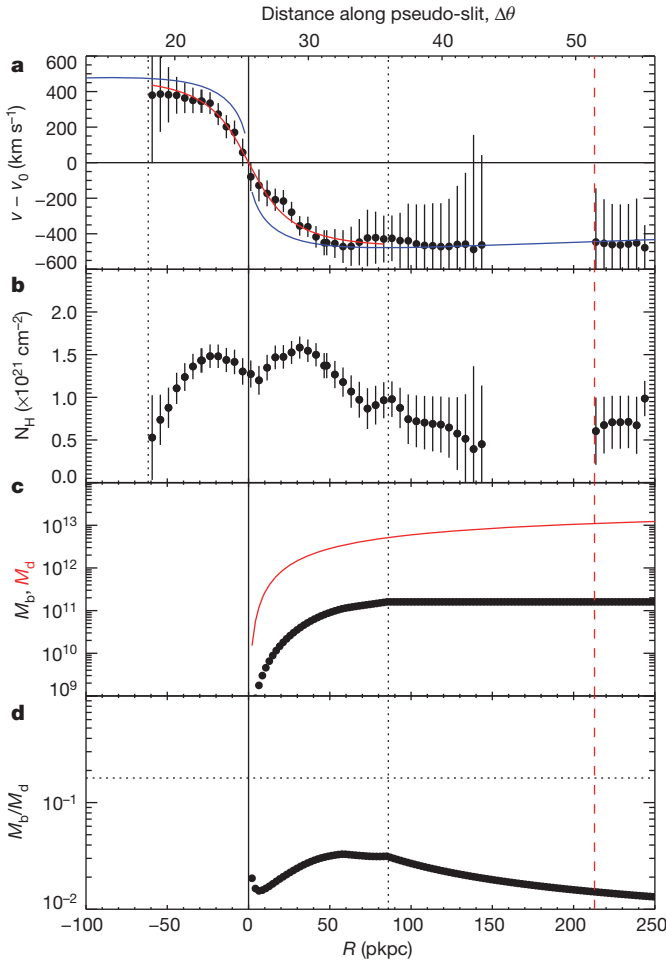


Figure 3 | Physical properties of the extended disk. **a**, Mean velocity (symbols) along the pseudo-slit (Fig. 2) with respect to the disk central velocity v_0 , with vertical lines showing an approximately 150 pkpc extent of the disk (dotted black lines), and virial radius (red dashed line). The extent of the filament profile is $80 \text{ pkpc} < R < 250 \text{ pkpc}$, where R is the radial coordinate of the disk. The NFW profile fits with and without PSF convolution are plotted in red and blue, respectively. Error bars are $\pm 1\sigma$. **b**, Inferred hydrogen column densities N_{H} assuming $t_{\text{d}} = 3 \text{ pkpc}$. Error bars are $\pm 1\sigma$ and do not include the uncertainty due to disk thickness. **c**, The cumulative baryonic mass M_{b} (dots) and dark-matter mass M_{d} (red line) within a sphere of radius R . **d**, Ratio of baryonic mass inside R to dark-matter mass (symbols) and the canonical ratio of baryonic to dark-matter mass of 0.17 (horizontal dotted line).

in the Lyman continuum, and that the column density is a strong function of disk thickness and flux only, as previously shown^{13,14}. The brightness of the fluorescent Ly α emission suggests that the dust-to-gas ratio in the disk is low (implying low metallicity), because a substantial reduction in output flux at a hydrogen column density $N_{\text{H}} \approx 10^{21} \text{ cm}^{-2}$ occurs at dust-to-gas ratios greater than 10^{-3} – 10^{-2} times the standard Milky Way ratio¹⁵.

Because the inferred gas mass is proportional to the square root of the disk thickness t_{d} , we require an estimate of t_{d} . For a self-gravitating gas disk supported by thermal or turbulent pressure, the thickness is related to the gas temperature and column density: $t_{\text{d}} \approx 0.35 \left(\frac{T_{\text{d}}}{10^4 \text{ K}} \right) \left(\frac{N_{\text{H}}}{10^{21} \text{ cm}^{-2}} \right)^{-1} \text{ pkpc}$. For all the models in our parameter grid, the disk temperature $T_{\text{d}} \approx 3 \times 10^4 \text{ K}$, which implies a disk thickness of at least 1 pkpc and a plausible disk gas density $n_{\text{H}} \approx 0.5 t_{\text{d}}^{-0.48} \text{ cm}^{-3}$. Simulations that produce extended disks⁴ suggest a thickness that is less than 10 pkpc. Such a disk would have three times the column density of the 1-pkpc case and require turbulent velocity support to maintain the 10-pkpc thickness of $\sigma_{\text{v}} \approx 60 \text{ km s}^{-1}$ (for a pressure-supported disk $t_{\text{d}} \approx 5 \left(\frac{\sigma_{\text{v}}}{50 \text{ km s}^{-1}} \right)^2 \left(\frac{N_{\text{H}}}{10^{21} \text{ cm}^{-2}} \right)^{-1}$), which is comparable to the velocity

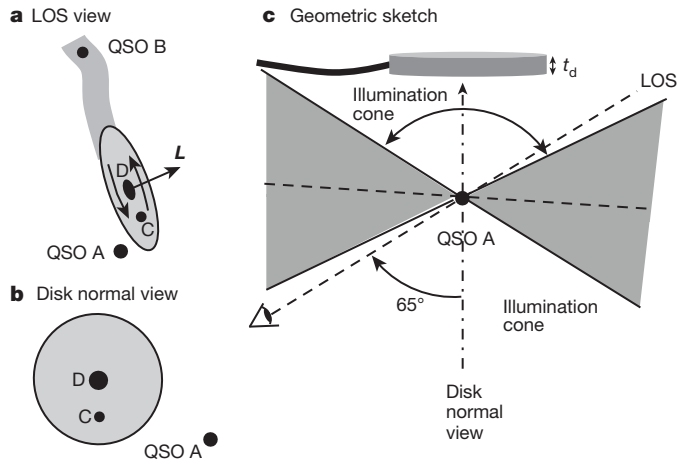


Figure 4 | Sketch of the geometry of the QSO-disk system. **a**, Line of sight (LOS) view as seen by the telescope/PCWI showing both QSOs, the disk, the filament, and the angular momentum vector for the disk, L . **b**, View normal to the disk, for the case of a 125-pkpc distance between QSO A and the disk. **c**, Approximate geometry (showing the QSO, the LOS, the QSO illumination cones (60° half-angle, $\pi \text{ sr}$ per cone), for the case of a 125-pkpc distance between QSO A and the disk.

dispersions in high-redshift disks and is less than the approximately 80 – 90 km s^{-1} range that would produce efficient shocking and radiative cooling. Additionally, the presence of a clear emission deficit in the central 25 pkpc suggests that the disk is much thinner than $25 \cos(70^\circ) \approx 8.6 \text{ pkpc}$. Consequently, we take 1–10 pkpc as a reasonable range and use $t_{\text{d}} = 3 \text{ pkpc}$ for all the mass estimates below, which have errors of $\pm 0.25 \text{ dex}$. We infer the column density as a function of disk radius, assuming a 3-pkpc-thick disk (Fig. 3b).

The estimates above allow us to compare the halo mass profile derived from the rotation fits to the inferred baryonic gas mass profile (Fig. 3c, d). We find that the total baryonic mass in the disk is $M_{\text{b}} = 10^{11.2} \left(\frac{t_{\text{d}}}{3 \text{ pkpc}} \right)^{0.52} M_{\odot}$, where M_{\odot} is the solar mass. The ratio of baryonic mass to total mass is $0.013 \left(\frac{t_{\text{d}}}{3 \text{ pkpc}} \right)^{0.52}$, which for $t_{\text{d}} = 3 \text{ pkpc}$ is about 8% of the canonical ratio 0.17. As expected, the baryonic mass is about 8% of the baryonic Tully–Fisher relation¹⁶ for $v_{\text{c}} = 470 \text{ km s}^{-1}$. The low baryon fraction in the cold disk component may imply that either the baryons are lagging the dark-matter collapse or that a substantial component is shocked to a much higher virial temperature. However, the mass profiles in Fig. 3c, d are quite similar, with only a small additional deficit in the inferred baryon ratio at the centre of the disk. We estimate that the filament has a mass $M_{\text{fil}} \approx 10^{11} M_{\odot}$, which is comparable to the disk mass, and suggests that some of the baryons associated with the halo will accrete in the future.

The integral field spectroscopic data allows us also to determine the total angular momentum of the disk baryons: $L_{\text{b}} \approx 3.2 \times 10^{15} t_{\text{d}}^{0.52} M_{\odot} \text{ kpc km s}^{-1}$. The specific angular momentum j_{b} is independent of the disk thickness: $j_{\text{b}} = L_{\text{b}}/M_{\text{b}} = 2 \times 10^4 \text{ pkpc km s}^{-1}$. These quantities, in conjunction with the halo parameters, allow us to determine the normalized spin parameter¹⁷: $\lambda_{\text{b}} = j_{\text{b}} / (\sqrt{2} R_{\text{vir}} v_{\text{vir}}) = 0.14$, where v_{vir} is the virial velocity. Typically, dark-matter halos have spin parameters $\lambda_{\text{d}} \approx 0.04$ at $z = 0$, although spin parameters are predicted to be higher at $z = 2$, sometimes reaching⁵ $\lambda_{\text{d}} \approx 0.1$. The baryonic spin parameter that we measure is an additional 40% larger, suggesting that the baryons in the extended disk have an angular momentum that is much greater than that of the dark matter halo. The large diameter and very high angular momentum of the structure strongly suggest that the extended disk is a ‘cold flow disk’, similar to those predicted by simulations⁴. This is the first directly imaged example of such an object, to our knowledge.

The extended disk of QSO UM287 does not have a detected stellar component in the continuum image, although there is continuum emission in the central part of the disk that coincides with the Ly α emission deficit (source D, another bright continuum emission object, is located almost exactly at the centre of the disk). The column density that we estimate, $N_{\text{H}} \approx 2.5 \times 10^{21} \text{ cm}^{-2}$, is above the star-formation threshold for solar metallicity. Although feedback from the QSO illumination could halt star formation for 1–10 Myr, the continuum rest-frame far-ultraviolet emission probes a 100-Myr timescale. The lack of star formation in the extended disk could be explained by a low metallicity ($Z < 0.1$), which raises the star-formation threshold, owing to the lower dust content and H₂-formation rate¹⁸. Source D implies a central star-formation rate of $>15 M_{\odot} \text{ yr}^{-1}$; the Ly α emission deficit could be caused either by increased dust absorption or by a gas deficit. The central deficit could also be produced by preferential photoevaporation by the QSO, or H I in the foreground (on the observer's side of the QSO) given that H I is abundant in the circumgalactic medium around quasars¹⁹.

Other possible causes of the observed velocity shear deserve comment, such as whether the extended disk is a result of the interactions that produced QSO source A and another bright object, QSO source B. We explore some alternatives in the Methods. QSO-merging disk models and observations imply, in general, a late appearance of the optical QSO phenomenon and only faint and fading signs of interactions. The large size, smooth kinematics and excellent fit to a simple disk model are not consistent with merging disks and tidal tails. The lack of any direct connection to QSO A or velocity shear nearing and crossing QSO B suggest that these objects are not directly linked or fed by the disk. The observations are best explained by an extended rotating disk linked to a cosmic web filament.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 February; accepted 21 May 2015.

Published online 5 August 2015.

1. Birnboim, Y. & Dekel, A. Virial shocks in galactic haloes? *Mon. Not. R. Astron. Soc.* **345**, 349–364 (2003).
2. Dekel, A. & Birnboim, Y. Galaxy bimodality due to cold flows and shock heating. *Mon. Not. R. Astron. Soc.* **368**, 2–20 (2006).
3. Kereš, D., Katz, N., Weinberg, D. H. & Davé, R. How do galaxies get their gas? *Mon. Not. R. Astron. Soc.* **363**, 2–28 (2005).

4. Stewart, K. *et al.* Orbiting circumgalactic gas as a signature of cosmological accretion. *Astrophys. J.* **738**, 39 (2011).
5. Stewart, K. R. *et al.* Angular momentum acquisition in galaxy halos. *Astrophys. J.* **769**, 74 (2013).
6. Cantalupo, S., Arrighi-Battaia, F., Prochaska, J. X., Hennawi, J. F. & Madau, P. A cosmic web filament revealed in Lyman- α emission around a luminous high-redshift quasar. *Nature* **506**, 63–66 (2014).
7. Matuszewski, M. *et al.* The Cosmic Web Imager: an integral field spectrograph for the Hale telescope at Palomar Observatory: instrument design and first results. In *Proc. SPIE, Ground-based and Airborne Instrumentation for Astronomy III* (eds McLean, I. S. *et al.*) 77350P (SPIE, 2010).
8. Martin, D. C. *et al.* Intergalactic medium emission observations with the Cosmic Web Imager. I. The circum-QSO medium of QSO 1549+19, and evidence for a filamentary gas inflow. *Astrophys. J.* **786**, 106 (2014).
9. Martin, D. C. *et al.* Intergalactic medium emission observations with the Cosmic Web Imager. II. Discovery of extended, kinematically linked emission around SSA22 Ly α blob 2. *Astrophys. J.* **786**, 107 (2014).
10. Prescott, M. K. M. *et al.* Resolving the galaxies within a giant Ly α nebula: witnessing the formation of a galaxy group? *Astrophys. J.* **752**, 86 (2012).
11. Prescott, M. K. M., Martin, C. L. & Dey, A. Spatially resolved gas kinematics within a Ly α nebula: evidence for large-scale rotation. *Astrophys. J.* **799**, 62 (2015).
12. Ferland, G. J. *et al.* The 2013 release of Cloudy. *Rev. Mex. Astron. Astrofis.* **49**, 137–163 (2013).
13. Bertoldi, F. The photoevaporation of interstellar clouds. I. Radiation-driven implosion. *Astrophys. J.* **346**, 735–755 (1989).
14. Hennawi, J. F. & Prochaska, J. X. Quasars probing quasars. IV. Joint constraints on the circumgalactic medium from absorption and emission. *Astrophys. J.* **766**, 58 (2013).
15. Neufeld, D. A. The transfer of resonance-line radiation in static astrophysical media. *Astrophys. J.* **350**, 216–241 (1990).
16. McGaugh, S. S., Schombert, J. M., de Blok, W. J. G. & Zagursky, M. J. The baryon content of cosmic structures. *Astrophys. J.* **708**, L14–L17 (2010).
17. Bullock, J. S. *et al.* A universal angular momentum profile for galactic halos. *Astrophys. J.* **555**, 240–257 (2001).
18. Krumholz, M. R., McKee, C. F. & Tumlinson, J. The star formation law in atomic and molecular gas. *Astrophys. J.* **699**, 850–856 (2009).
19. Prochaska, J. X. *et al.* Quasars probing quasars. VI. Excess H I absorption within one proper Mpc of $z \sim 2$ quasars. *Astrophys. J.* **776**, 136 (2013).

Acknowledgements We thank T. Tombrello and S. Kulkarni for their support of PCWI. This work was supported by the National Science Foundation and the California Institute of Technology.

Author Contributions D.C.M. is the principal investigator of PCWI, led the observations and analysis of UM287, and was principal author on the paper. M.M., D.C., and P.M. designed, constructed, and operated PCWI. A.M. was the project and technical manager (2006–2010). J.D.N., M.M., and D.C.M. developed the PCWI/KCWI data pipeline and produced the final data cubes. M.M., P.M., S.C., and J.X.P. contributed Keck data and helped edit the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.C.M. (cmartin@srl.caltech.edu).

METHODS

Instrument and observations. We have constructed an integral field spectrograph, called the Palomar Cosmic Web Imager (PCWI), that is designed to search for, map, and characterize intergalactic-medium emission and other low surface brightness phenomena⁷. It is built with a $40'' \times 60''$ reflective image slicer with twenty-four $40'' \times 2.5''$ slices. PCWI is mounted at the Cassegrain focus of the Hale 5-m telescope on Mt Palomar, USA. The imaging resolution, while limited by the $2.5''$ slicer sampling, can be effectively improved to approximately $1.3''$ by dithering the field between individual exposures. A description of the instrument, general observing approach, and data analysis methodology is given in refs 8 and 9. For the UM287 observation, the spectrograph was fitted with a Richardson reflection grating, blazed near $5,000 \text{ \AA}$, and has an instantaneous bandwidth of approximately 170 \AA with the nod-and-shuffle mask in place. With this grating, the spectrograph attains a slit-limited resolution $\Delta\lambda \approx 2.5 \text{ \AA}$ and a peak efficiency of about 4% at $4,000 \text{ \AA}$ including the telescope and atmosphere.

We obtained a total of 2 h on-source and 2 h off-source exposure centred on QSO UM287 on 23 September 2014. Individual exposures were acquired using the nod-and-shuffle technique^{20–22}. The PCWI implementation of nod-and-shuffle uses the central third of the charge-coupled device (CCD) to record the spectrum, and masks the outer two thirds of the detector for storage (restricting the effective common bandpass to about 170 \AA given the image slicer offset brick wall pattern and slit curvature). Individual exposure frames are created by interleaving an integral number, N , of on-target telescope pointings of length t with $N + 1$ background pointings, the first and last of length $t/2$, and the remainder of length t . This cadence results in separate source and background tiles being built up on the CCD, each equivalent to an $(N \times t)$ exposure. A typical 40-min exposure ($t = 120 \text{ s}$, $N = 10$; 20 min source and 20 min background) takes approximately 50 min of wall-clock time, including CCD read-out. The benefits of using the nod-and-shuffle method are that the sky is sampled frequently and nearly contemporaneously with the object, it is imaged through the same optical path as the object, and it is recorded using the same detector pixels as the object. This improves sky subtraction precision, reduces the contribution of detector read noise by decreasing the number of CCD read-outs, and limits the impact of instrument systematics. Pixel binning the detector 2 by 2 further reduces the impact of the three-electron (root mean square) read noise.

Data cube generation. We collected numerous calibration images using internal and dome lamps throughout the observing period. The obtained 2D spectra were processed with the PCWI/Keck CWI data reduction pipeline. They were rectified and aligned using geometric mask calibration images and arc lamp spectra. Background panels were subtracted from source panels. Individual 3D data cubes were then wavelength-shifted to compensate for the $<1 \text{ \AA}$ of total flexure using sky lines. The final mosaicked and co-added data cubes were assembled using the astrometry on the basis of the QSO position. The reconstruction is accurate to about 0.5 arcsec (root mean square).

Exposure maps were generated by processing normalized flat-field images (calibration and twilight flats) in a similar fashion. The result is a set of data cubes (RA, dec., and λ) for each exposure, sampled at $(0.55'', 1.1 \text{ \AA})$ and covering $3,940\text{--}4,110 \text{ \AA}$. As the nod-and-shuffle mask does not physically contact the CCD, a small amount of diffuse continuum light scatters underneath it and remains in the subtracted cube ($<1\%$). This residual is easily subtracted with a low-order continuum fit. The data cube was normalized using the measured signal from the QSO UM287 and its measured spectrum at $4,000 \text{ \AA}$ from the Sloan Digital Sky Survey. The absolute flux measurements are accurate to $\pm 10\%$.

The data cube projections (other than Extended Data Fig. 1a–d) are produced from this difference flux cube by a 3D adaptive-smoothing-in- λ algorithm, which incorporates a hierarchical adaptive-smoothing algorithm in space and wavelength^{8,9}. The 3σ noise threshold used for this algorithm is derived directly from the difference cube, and is consistent with the predicted Poisson noise. Raw spectral and narrow-band images compare well with the smoothed images (Extended Data Fig. 1, Fig. 1).

Channel maps. As the filament approaches QSO UM287 (Fig. 1) the mean velocity shears red-ward and the dispersion broadens. A succession of 2-\AA channel cuts through the smoothed data cube (Extended Data Fig. 2) illustrate a disk morphology with emission moving linearly towards the QSO (but offset) as the channels move from blue to red. These images also demonstrate that neither the QSO subtraction nor the presence of a nearby line and continuum object (source C) are responsible for the disk emission nearest QSO UM287. To reveal the spatial structure of the emission, we use the kinematic and spatial behaviour gleaned from the channel maps (Extended Data Fig. 2) to design a new narrow-band window that follows the velocity-shear structure. Using this technique we find evidence of an extended gaseous disk offset from, and illuminated by, the nearby QSO (Fig. 3).

Bright QSO subtraction. We generated several figures in the text (Fig. 3, Extended Data Figs 2 and 5) by subtracting an average image. In the case of

spectral images, we calculated and then subtracted an average intensity versus slit position map. That is, we calculated a single flux at each slit position and then subtracted it from the spectral image—this can be thought of as a single-value continuum subtraction. In the case of narrow-band images, a single PSF is calculated by averaging over a wavelength range. This average image is subtracted from each narrow-band image. The wavelength range used is $3,958\text{--}4,013 \text{ \AA}$, and includes the Ly α emission region (we include this to be conservative since it subtracts a small amount of extended emission, but we note it does not noticeably alter the results). In Extended Data Fig. 2, we subtract only the emission within an elliptical contour (reflecting the PSF) near the bright QSO. In Fig. 2, we subtract the average over the full spectral image in order to remove both QSOs and highlight the disk emission. For the narrow-band images in Extended Data Fig. 2 and Fig. 2, we calculate an average image for the bright QSO and subtract that from the narrow-band image, to highlight the disk emission.

Nearby sources and source-subtraction residuals. It is important to demonstrate that the disk emission is not an artefact produced by the nearby QSO or the continuum source C. We show (Extended Data Fig. 3) the CWI, W. M. Keck Observatory narrow-band, and Keck V-band continuum image, along with locations of the QSOs (A and B) and two sources near QSO A (C and D). Source C is a fairly bright, compact continuum and line source near the QSO. Source D is a fainter, slightly extended continuum and Ly α line source also near the QSO. Sources C and D appear distinct in both the Keck continuum and the narrow-band images. Source D falls exactly in the disk emission minimum and centre. Source C is offset north from the centre of the disk (Fig. 2). Residuals following the elliptical contour are outside the subtraction area. When we perform no subtraction (Extended Data Fig. 4), we see that the excess emission associated with the disk and source C are still present in all channels, and in particular in the positive velocity channels. We can also remove the average emission from each entire image, not just near QSO A (Extended Data Fig. 5). Again the average is taken over the range $3,958\text{--}4,013 \text{ \AA}$ including the Ly α emission region. Source C is removed because it is a continuum source, except in the channels in which its line emission appears (Extended Data Fig. 5j, k). Considerable positive-velocity disk emission remains after removing the continuum, particularly in Extended Data Fig. 5h–j, k, and clearly exceeds any small remaining residuals due to small (around 1%) variations of the PSF with wavelength. This emission is well offset from source C. Finally, we can use a stepped vertical pseudo-slit to show a kinematically sheared profile that is consistent with a disk at several locations on and off QSO A (Extended Data Fig. 6).

Kinematic modelling. We model the 2D velocity profile (Fig. 2c, e) with a Navarro–Frenk–White (NFW) dark-matter halo profile²³. We assume that the velocity profile is circular and dominated by dark matter with no stellar or baryonic contribution. We convolve the predicted 2D profile with the measured PSF based on the QSO A image. The 2D velocity profile is well fitted (Fig. 2f, g) by a rotating disk in an NFW dark-matter halo with mass $\log_{10} M_h = 13.1^{+0.6}_{-0.3}$, halo concentration $c = 3^{+5}_{-2}$, central velocity near systemic ($v_0 = 50 \pm 30 \text{ km s}^{-1}$), circular velocity (at the virial radius) $v_c \approx 500 \text{ km s}^{-1}$, and virial radius $R_{\text{vir}} = 225 \text{ pkpc}$. Using a pseudo-slit (Fig. 2) we extracted a 1D velocity profile (Fig. 3a), which is also well fitted by an NFW model with consistent inferred geometry, mass $\log_{10} M_h = 13.1^{+0.5}_{-0.3}$, and halo concentration $c = 5^{+7}_{-3}$. The blue-shifted end of the disk ($v \approx -450 \text{ km s}^{-1}$) smoothly transitions to an extended filament ($80 \text{ pkpc} < R < 250 \text{ pkpc}$, crossing the faint QSO) with the same velocity of about -450 km s^{-1} .

Intensity modelling. To constrain the geometry of the disk with respect to the LOS and the QSO, we set up a simple model (Fig. 4). We assume that the QSO has two opposite emission cones each with $\pi \text{ sr}$ (60° cone half-angle). The line of sight is included in one of these cones, and the disk and the filament feeding the disk are included in the other. The disk inclination is estimated from the axis ratio and the 2D kinematic model fit to be about $70^\circ \pm 5^\circ$. The requirement that the rear emission cone illuminates all of the visible disk and filament places a lower limit on the separation between the disk and the QSO of 65 pkpc . We note that QSO B could also produce some of the illumination, but we neglect that here.

We use the ‘Cloudy’¹² nebular emission code (version 11.10) to estimate the column density and constrain the gas density of the illuminated disk. We use a standard QSO spectrum with luminosity $\log_{10}(vL_\nu) = 46.9$, black-body temperature $T = 1.5 \times 10^5 \text{ K}$, and power-law indices $\alpha_{\text{ox}} = -1.4$ (for optical-to-X-ray), $\alpha_{\text{UV}} = -0.5$ (for UV), and $\alpha_x = -1.0$ (for X-ray). We populate a grid of fixed number density models with total gas column density $\log_{10} N_H = \{20.5, 20.6, \dots, 22.5\}$, distance between the disk and QSO $A r_0 = \{75, 125, \dots, 275\} \text{ pkpc}$, disk thickness $t_d = \{1, 3, 10, 30\} \text{ pkpc}$, disk turbulent velocity $\sigma_v = \{0, 300\} \text{ km s}^{-1}$, disk metallicity $[Z/H] = \{0, -1, -2\}$. For each of these models a Ly α flux is calculated, with contributions from recombination radiation, line scattering, and collisional excitation also determined. We find that the model predicts that intensity only strongly depends on disk thickness (disk density) and hydrogen column

density, as expected, since the neutral fraction ($n_{\text{H}^0}/n_{\text{H}}$) scales as the square root of the gas density. The model predicts that the gas is optically thin in the Lyman continuum, and thus column density only a strongly depends on disk thickness and flux, as previously shown^{13,14}. For recombination radiation we expect $N_{\text{H}} \approx (It_{\text{d}})^{0.5} \text{ cm}^{-2}$ where I is the intensity; the resulting model nearly replicates this dependence: $N_{\text{H}} = 10^{21.15} t_{\text{d}}^{0.52} f_{0.56}^{0.09} R_{100}^{0.07} \sigma_{\text{v}}^{0.06} [Z/H]^{-0.006}$, where f_0 is the Ly α flux in units of $3 \times 10^{-17} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ arcsec}^{-2}$ and R_{100} is the distance between the QSO and the disk in units of 100 kpc. The neutral hydrogen column density N_{H} is relatively independent of disk thickness, and varies strongly with QSO/disk separation, ranging from $N_{\text{H}} = 10^{16.2} \text{ cm}^{-2}$ for $R_{100} = 0.75$ to $N_{\text{H}} = 10^{17.0} \text{ cm}^{-2}$ for $R_{100} = 2.0$. In the regime determined from the data, the neutral fraction is low ($\ll 1$) and most of the Ly α emission is produced by radiative recombination (as opposed to line scattering^{8,24} or collisional excitation²⁵). The model is insensitive to whether the disk is thermally or turbulently supported, because, in the model grid, the thickness and column density are fixed to enable us to derive the intensity, and the line emissivity is only a weak function of the turbulent velocity σ_{v} .

The brightness of the fluorescent Ly α suggests that the dust-to-gas ratio in the disk is low, because substantial reduction in output flux at $N_{\text{H}} \approx 10^{21} \text{ cm}^{-2}$ occurs with dust-to-gas ratios greater than 10^{-3} – 10^{-2} times the standard Milky Way ratio¹⁵. Finally, we use the average intensity at each point of the pseudo-slit in Fig. 2 to determine N_{H} as a function of the other parameters. We estimate the filament mass using the same model. Here, the filament thickness is taken to be 24 pkpc, on the basis of a filament width of approximately 3 arcsec, length of 160 pkpc, and a mean $N_{\text{H}} \approx 7 \times 10^{20} \text{ cm}^{-2}$ (Fig. 3). The filament mass is then $M_{\text{fil}} \approx 10^{11.52} M_{\odot}$, where $t_{\text{fil}} = 24 \text{ pkpc}$.

We compute the angular momentum of the disk by assuming that the fitted velocity profile (intrinsic, before blurring and averaging) and column density profile apply to the full 180° associated with each half of the disk.

Radiative transfer. Optically thick Ly α radiation can produce a double-peaked profile as line photons escape from the line centre by frequency random walk¹⁵. The optical depth τ_0 for an H I column density $N_{\text{H}} \approx 10^{17} \text{ cm}^{-2}$ is $\tau_0 \approx 3,000$, which for a gas temperature $T \approx 10^4$ – 10^5 K (and equivalent turbulent velocity) gives a double-peak separation less than 1.5 \AA . This separation would not be resolved in our observation. Under these conditions, for line photons escaping a relatively kinematically quiet disk ($\sigma_{\text{v}} < 50 \text{ km s}^{-1}$), and because we do not expect substantial velocity shears along the line of sight, the line centroid should reflect the average local gas velocity^{10,11}.

Alternative models. Since the UM287 disk is found in proximity to QSO A (and B), it is possible that it was produced by the same interactions that produced QSO A (and possibly B), or in interactions by the halo that the object could share with QSO A (and possibly B). Simulations of tidal tails in cold dark matter models suggest that long tails can be formed of about 250 pkpc^{26–28}. Here we make some general observations, and in this discussion refer to the UM287 nebula as being composed of the bright filament (Fig. 2, at station 20–35 arcsec) and the faint filament (Fig. 2, station 35–50 arcsec). Luminous QSOs at high redshift are probably formed from major mergers of two gas-rich disks²⁹. The optical QSO phase is late in the evolutionary sequence of such a merger and tends to show only faded and faint evidence of interactions such as tidal tails^{29–34}. Our kinematic observations of the UM287 nebula are well fitted by a smooth rotating disk connected to a quiescent filament with low velocity shear. The intensity distribution is also disk-like. The inferred gas column densities ($10^{21.5} \text{ cm}^{-2}$) and gas masses ($10^{11} M_{\odot}$) are high. The span of the bright filament is 120 pkpc and of the entire filament is several hundred proper kiloparsecs.

We consider three classes of alternative models. First, an undetected disk could be interacting with the host of QSO B, leading to a merging disk and long tidal tails illuminated by QSO A. However, there is no evidence in Ly α for emission or the complex kinematics of a merging disk at, or near, QSO B. Long tidal tails are usually fairly thin, curved, fade away from the interaction region, and show continuous velocity shear as various parts of the tail expand, rotate, and fall back³⁵. Neither the intensity, morphology, nor velocity profiles of the UM287 filament exhibit this behaviour. No emission is seen from an opposing tail.

A second possibility is that the bright part of the filament is part of a merging disk hosting QSO A or the tidal tail, and that the faint filament is part of the tidal

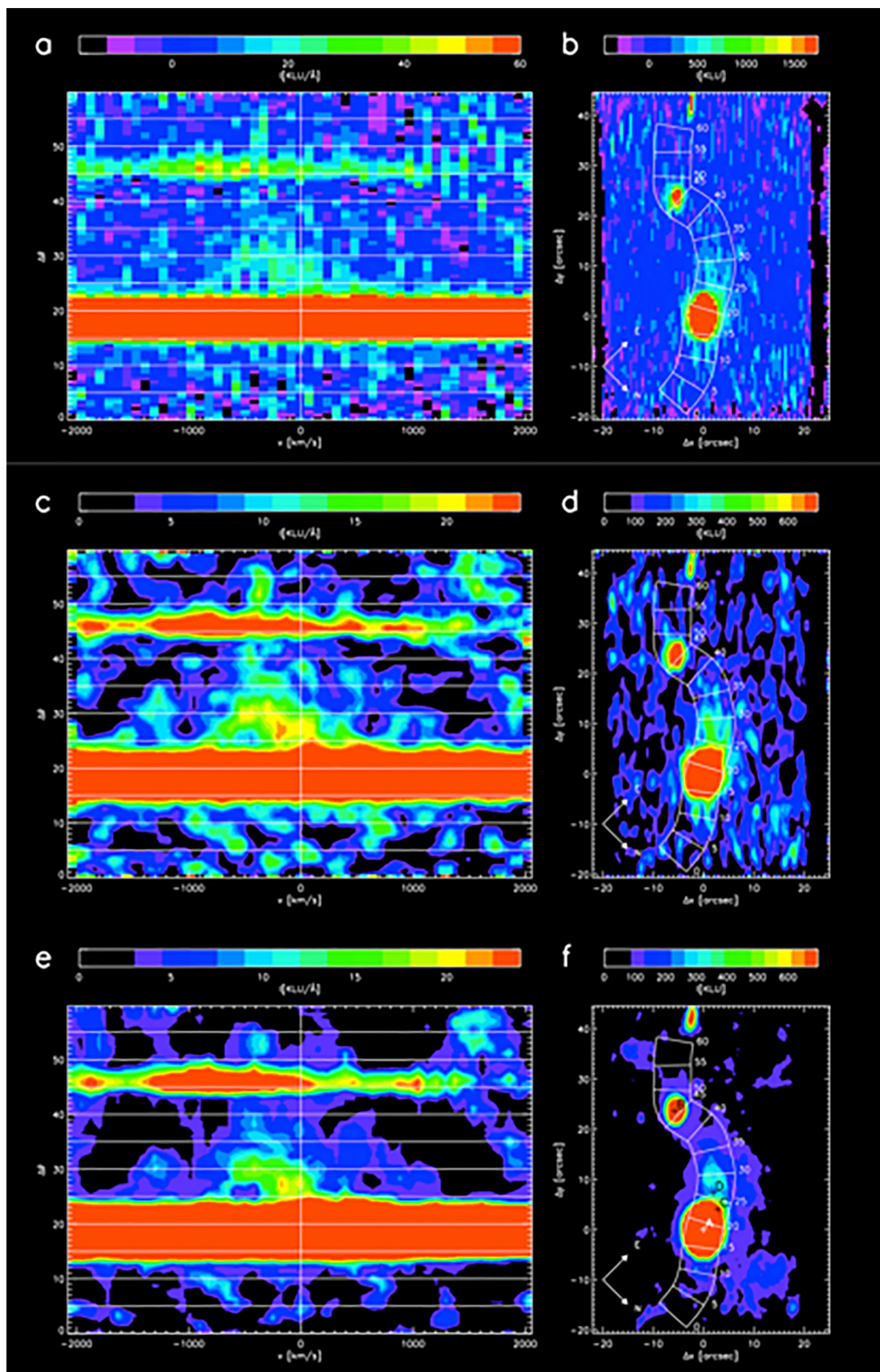
tail. Several issues arise with this picture. The bright filament is tangential to QSO A with no detected connection, whereas a tidal tail would be expected to project radially outward and then curve tangentially, all in a plane that includes the host merging disk and QSO A. It is difficult to illuminate the bright filament with QSO A unless their separation is at least 100 pkpc, which is inconsistent with the possibility of the bright filament being part of the merging disks. There is no evidence for complex merging disk/tail kinematics, for example large velocity shears where the tidal tail meets the merging disks, or curvature in velocity-position space of the tail³⁵. There are no continuum counterparts (other than D) to the filament, as usually seen in tidal tails. The filament is quite wide (approximately 60 pkpc) compared to observed and modelled tails. Gas moving outward in a tidal tail would be likely to show a kinematic discontinuity with the outer disk rotation curve. Typically, the base of the tail is closer to the systemic velocity of the merging system (as well as the end, with a noticeable curvature in between), because this gas (farthest from the centre of mass at perigalacticon) has suffered a tidal impulse that launches the tidal tail. Although there is no direct way to determine whether the gas in the filament is in-falling from the cosmic web or moving outward in a tidal tail, the lack of this kinematic signature is further (indirect) evidence that the gas is flowing in and smoothly merging with the outer parts of the disk.

A third possibility is that the filament is produced by an interaction that does not produce the QSOs but occurs in the same halo. In this case, the main purpose of positing the interaction is to explain the large size of the filament. Again, the simple and smooth kinematics, the large width, the high gas mass, and the disk-like morphology suggest that this is neither a merging disk system or a tidally disrupted disk passing through the halo of QSO A.

Thus, although we cannot investigate all possible interaction geometries and scenarios here, a merger/tidal scenario is not favoured by our observations.

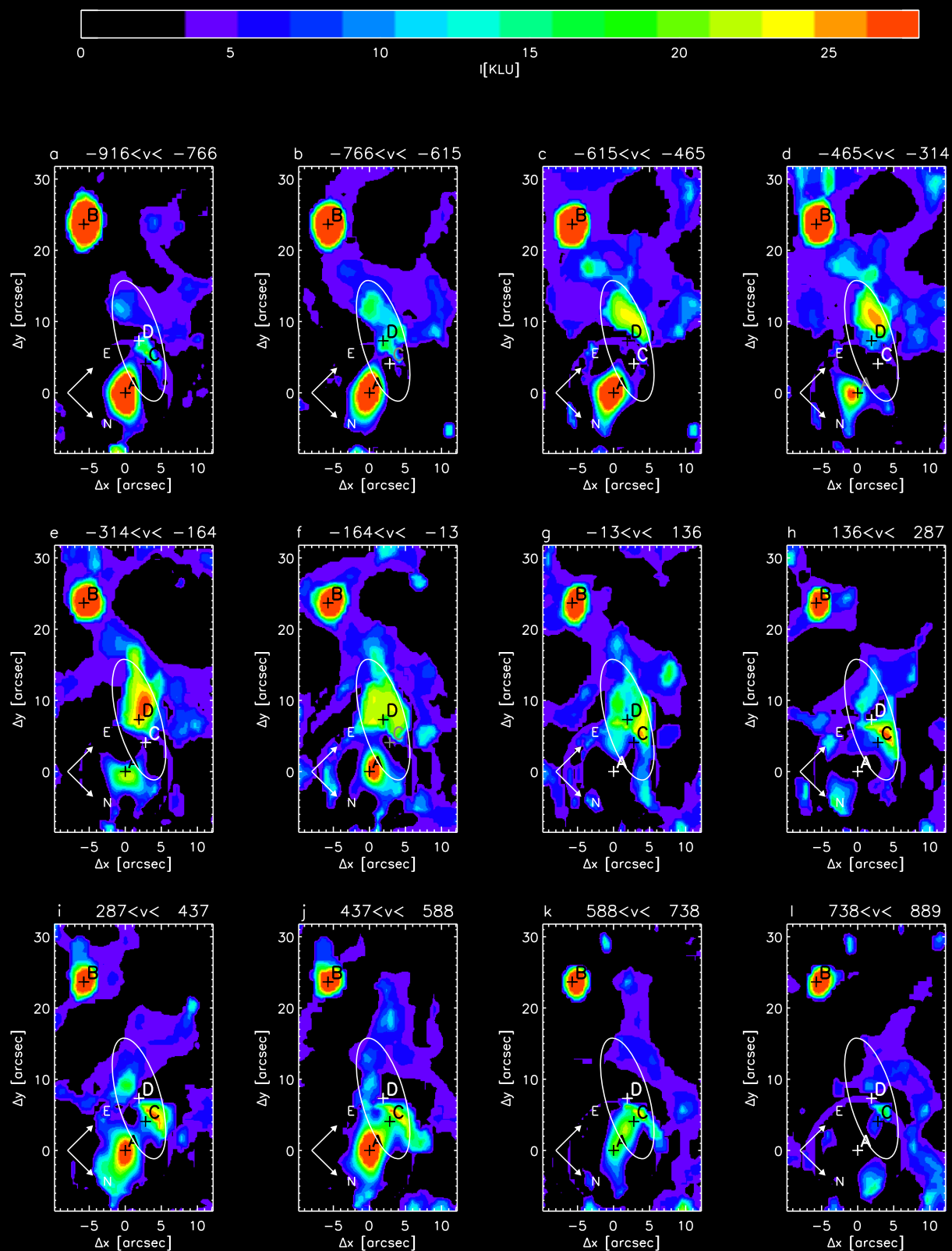
Code availability. We choose not to make the pipeline code available at this time because it is not fully documented for public use but plan to do so in early 2016.

20. Sembach, K. R. & Tonry, J. L. Accurate sky subtraction of long-slit spectra: velocity dispersions at $\Sigma_V = 24.0 \text{ mag/arcsec}^2$. *Astron. J.* **112**, 797–805 (1996).
21. Glazebrook, K. & Bland-Hawthorn, J. Microslit nod-shuffle spectroscopy: a technique for achieving very high densities of spectra. *Publ. Astron. Soc. Pacif.* **113**, 197–214 (2001).
22. Cuillandre, J. C. et al. “Va-et-Vient” spectroscopy: a new mode for faint object CCD spectroscopy with very large telescopes. *Astron. Astrophys.* **281**, 603–612 (1994).
23. Navarro, J. F., Frenk, C. S. & White, S. D. M. A universal density profile from hierarchical clustering. *Astrophys. J.* **490**, 493–508 (1997).
24. Cantalupo, S., Porciani, C., Lilly, S. J. & Miniati, F. Fluorescent Ly α emission from the high-redshift intergalactic medium. *Astrophys. J.* **628**, 61–75 (2005).
25. Cantalupo, S., Porciani, C. & Lilly, S. J. Mapping neutral hydrogen during reionization with the Ly α emission from quasar ionization fronts. *Astrophys. J.* **672**, 48–58 (2008).
26. Barnes, J. E. Encounters of disk/halo galaxies. *Astrophys. J.* **331**, 699–717 (1988).
27. Springel, V. & White, S. D. M. Tidal tails pin cold dark matter cosmologies. *Mon. Not. R. Astron. Soc.* **307**, 162–178 (1999).
28. Toomre, A. & Toomre, J. Galactic bridges and tails. *Astrophys. J.* **178**, 623–666 (1972).
29. Hopkins, P. F., Hernquist, L., Cox, T. J. & Kereš, D. A cosmological framework for the co-evolution of quasars, supermassive black holes, and elliptical galaxies. I. Galaxy mergers and quasar activity. *Astrophys. J.* **175** (Suppl.), 356–389 (2008).
30. Guyon, O., Sanders, D. B. & Stockton, A. Near-infrared adaptive optics imaging of QSO host galaxies. *Astrophys. J.* **166** (Suppl.), 89–127 (2006).
31. Hutchings, J. B. Host galaxies of $z \sim 4.7$ quasars. *Astron. J.* **125**, 1053–1059 (2003).
32. Hutchings, J. B., Cherniawsky, A., Cutri, R. M. & Nelson, B. O. Host galaxies of two micron all sky survey-selected QSOs at redshift over 0.3. *Astron. J.* **131**, 680–685 (2006).
33. Kawakatu, N., Anabuki, N., Nagao, T., Umemura, M. & Nakagawa, T. Type I ultraluminous infrared galaxies: transition stage from ULIRGs to QSOs. *Astrophys. J.* **637**, 104–113 (2006).
34. Urrutia, T., Lacy, M. & Becker, R. H. Evidence for quasar activity triggered by galaxy mergers in HST observations of dust-reddened quasars. *Astrophys. J.* **674**, 80–96 (2008).
35. Hibbard, J. E., van der Hulst, J. M., Barnes, J. E. & Rich, R. M. High-resolution H I mapping of NGC 4038/39 (“The Antennae”) and its tidal dwarf galaxy candidates. *Astron. J.* **122**, 2969–2992 (2001).



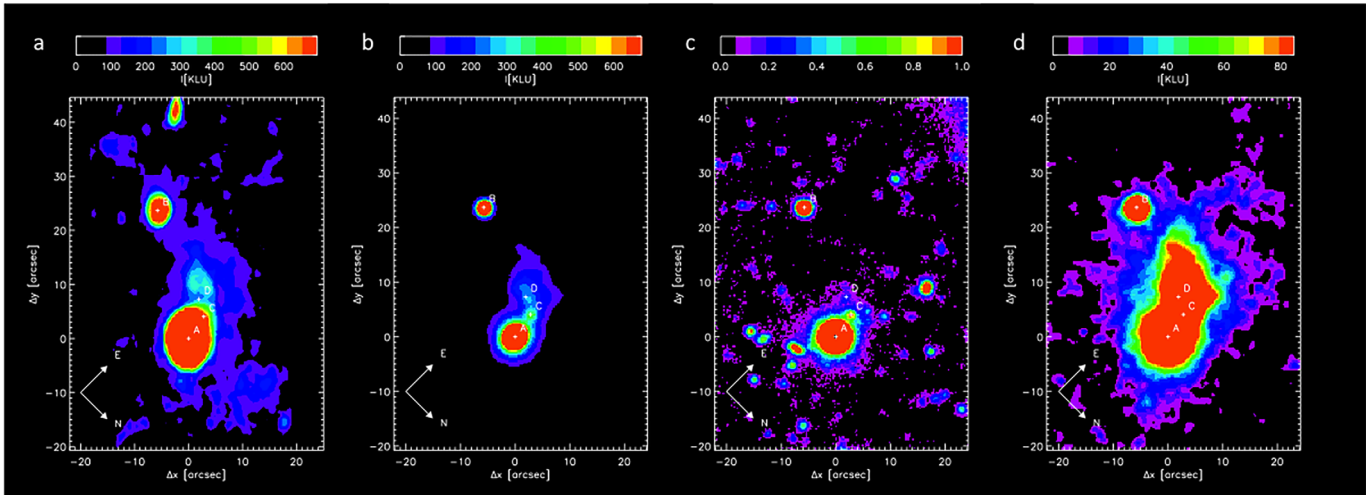
Extended Data Figure 1 | Illustration of raw and conventionally smoothed data. **a**, Raw-data spectral image of pseudo-slit obtained in the slit shown in **b**. 1σ error is about 3 kLU or approximately one colour scale step. **b**, Raw-data narrow-band image obtained in the 3,970–4,000 Å band. 1σ error is 50 kLU or about 0.5 colour scale steps. **c**, Raw spectral image shown in

a, conventionally boxcar smoothed by 10 pixels. **d**, Raw narrow-band image shown in **b**, conventionally boxcar smoothed by 10 pixels. **e**, Spectral image obtained by 3D adaptive smoothing as discussed in the Methods. **f**, Narrow-band image (3,970–4,000 Å) obtained by summing over the 3D adaptively smoothed data cube, as discussed in the Methods.



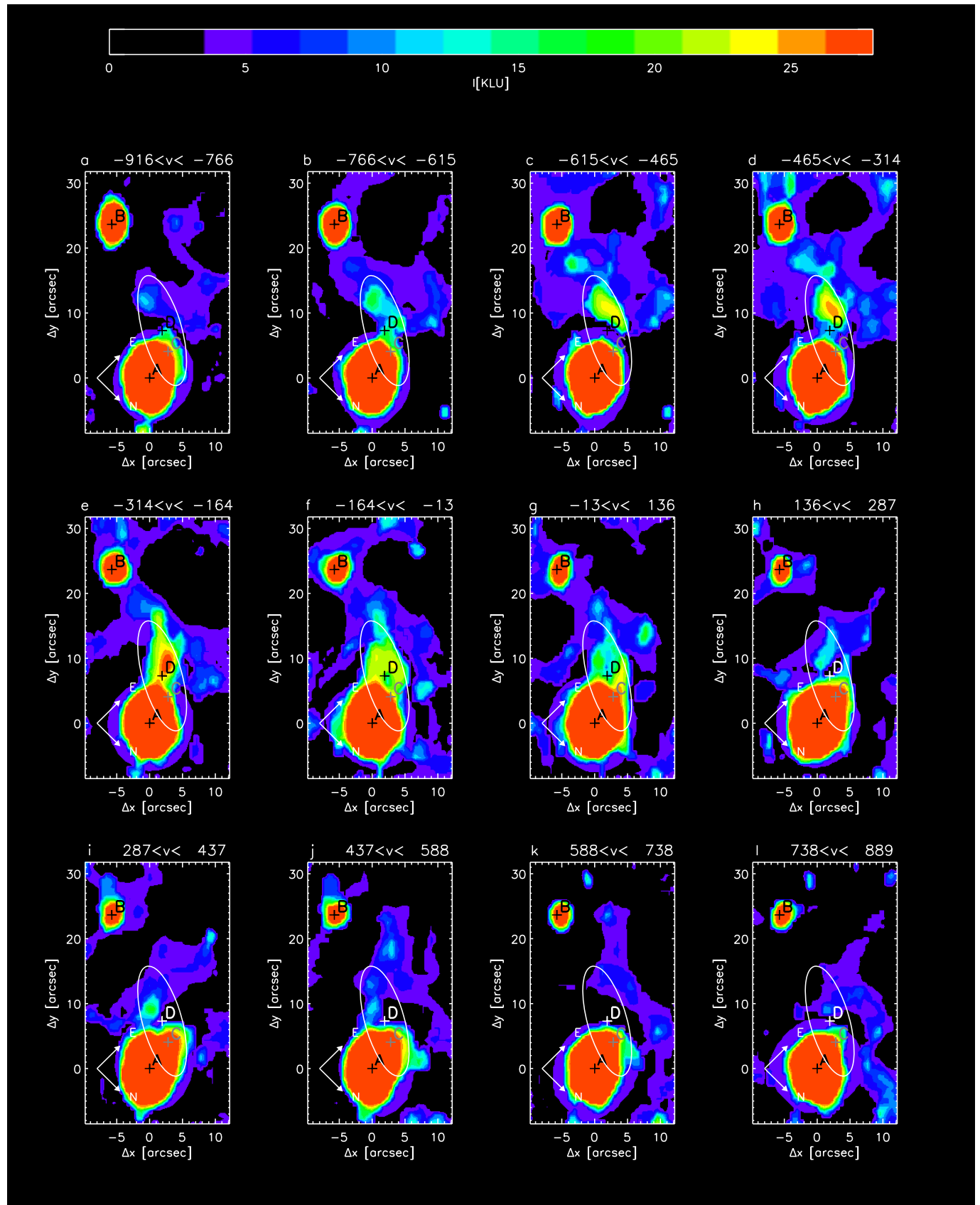
Extended Data Figure 2 | Channel maps of the UM287 data cube. **a–l**, Panels show individual velocity channels that are 150 km s^{-1} wide, corresponding to a 2 \AA width. Sources A–D near QSO UM287 are plotted. Velocities are with respect to the UM287 systemic velocity. QSO A has been subtracted by calculating an average PSF for the QSO over the $3,970\text{--}4,000 \text{ \AA}$ band, and then subtracting this slice by slice, within an elliptical radius of 6 arcsec in the x direction and 7.2 arcsec in the y direction. The residual flux in certain channels is due to three effects. (1) At the centre of QSO A, the residual flux is due to emission lines in the QSO around $\text{Ly}\alpha$. (2) At the elliptical boundary surrounding the QSO, a small subtraction residual can be seen outside the ellipse within which the subtraction is performed. This is typically 3–5 kLU, and can be seen clearly without additional sources in **c** and **d**. (3) Emission sources are present near the QSO in certain channels. Source C shows line emission primarily in **j** and **k** ($430 \text{ km s}^{-1} < v < 731 \text{ km s}^{-1}$), and its continuum

emission is subtracted along with the QSO. Any emission near the subtraction boundary above 5 kLU is not a subtraction residual due to the QSO. The disk emission appears bright in **b** southeast of QSO A, and moves north, approaching the QSO as the velocity moves redward. In the two central velocity channels, the emission appears almost ring-like and continues to move north, roughly centred on source D. The emission continues to move north in **h–j**. The emission is 15–25 kLU and therefore cannot be QSO subtraction residuals. The emission is also not partially subtracted emission from source C. The emission in **i** and **j** are several arcseconds east of source C. Further evidence for this point is given in Extended Data Figs 4 and 5. To indicate the approximate disk location, we show an elliptical contour with a major axis radius of 8.5 arcsec, position angle of 15.5° , and an ellipticity corresponding to an inclination of 70° .



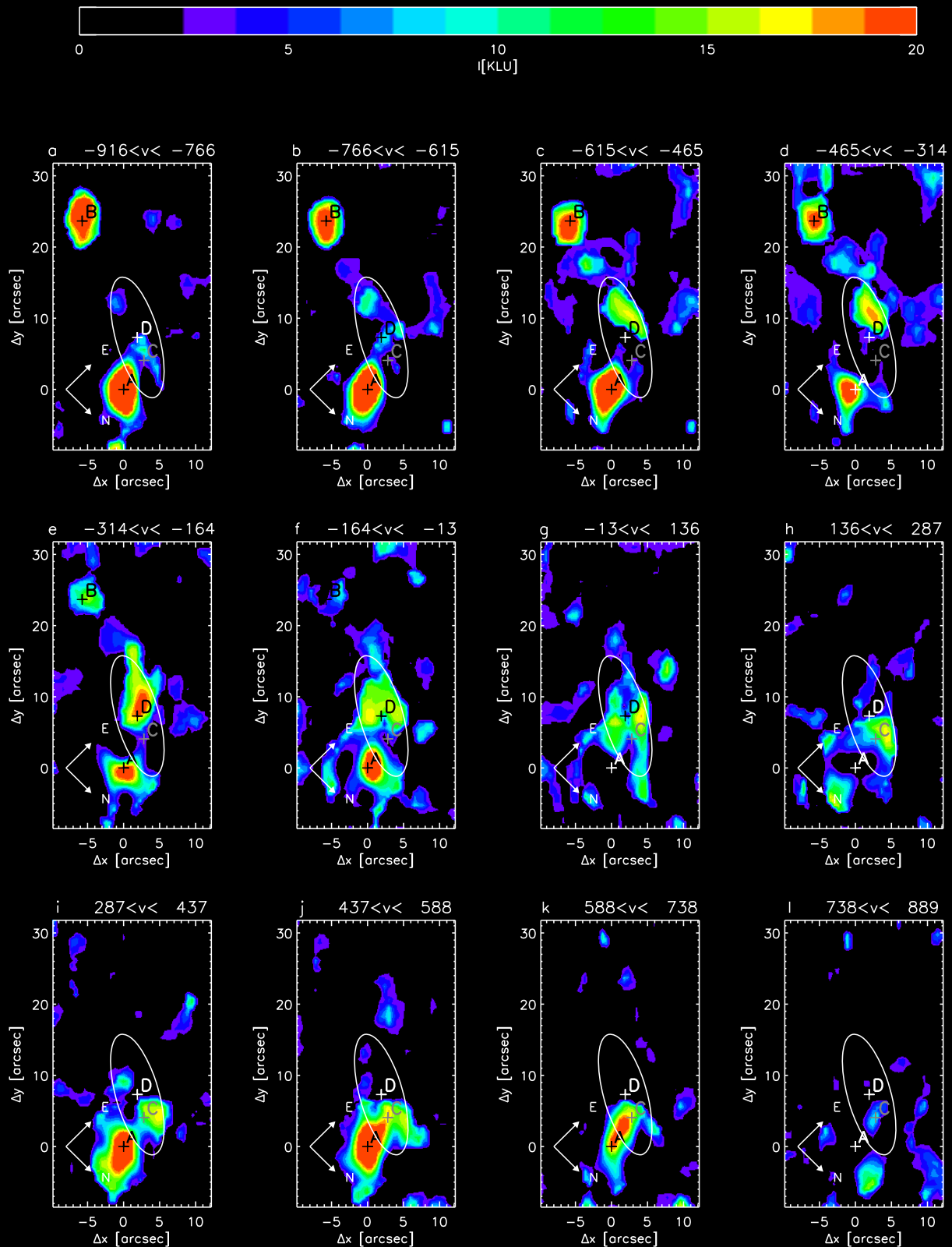
Extended Data Figure 3 | Comparison of PCWI data and Keck narrow-band and continuum images. **a**, PCWI narrow-band image created by summing 3,970–4,000 Å data-cube slices from the adaptively smoothed data cube. Sources A, B, C, and D are shown. PCWI image is not continuum subtracted. **b**, Keck continuum-subtracted narrow-band image on the same

intensity scale as the PCWI image in **a**. **c**, Keck V-band image. Sources A, C, and D are shown; continuum magnitudes in the V band are approximately (± 2 mag) 16.6 AB, 22.2 AB, and 23.8 AB respectively. **d**, Keck continuum-subtracted narrow-band image on an expanded intensity scale.



Extended Data Figure 4 | Channel maps of the UM287 data cube. Panels show individual velocity channels that are 150 km s^{-1} wide, corresponding to a 2 \AA width, as in Extended Data Fig. 2. In this case, no source subtraction has

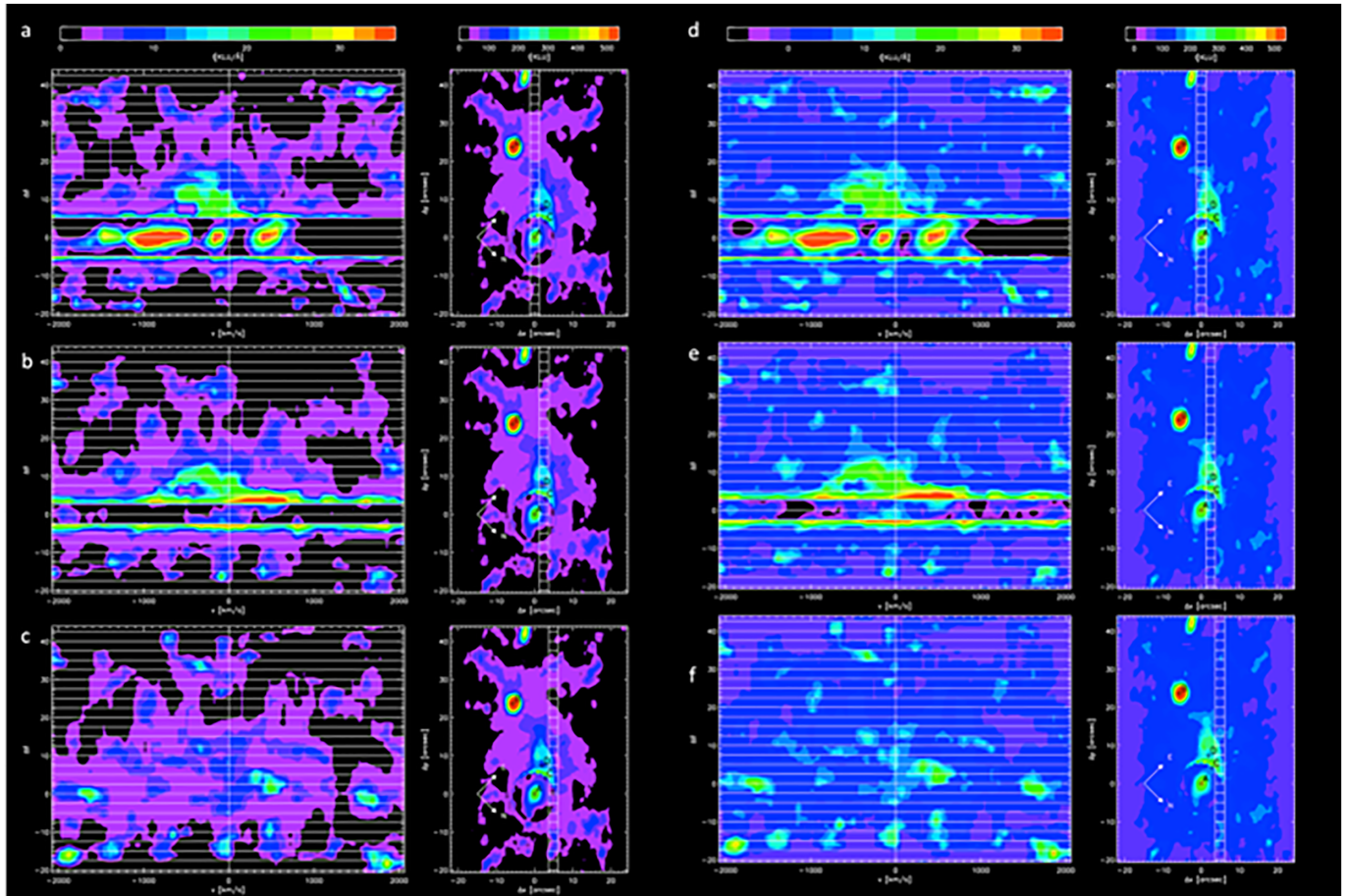
been performed. Sources near the QSO are plotted. Ellipses are drawn as in Extended Data Fig. 2. Velocities are with respect to the UM287 systemic velocity.



Extended Data Figure 5 | Channel maps of the UM287 data cube.

a–l, Individual velocity channels that are 150 km s^{-1} wide, corresponding to a 2 \AA width, as in Extended Data Fig. 2. In this case, the wavelength-averaged cube has been subtracted from the cube over the full field of view. Sources near the QSO are plotted. Ellipses are drawn as in Extended Data Fig. 2. Velocities

are with respect to the UM287 systemic velocity. This subtraction removes the average continuum from all sources, including source C. The emission centred on source C is present in **j** and **k** from line emission (presumably Ly α). The progression of the disk emission can be seen as in Extended Data Fig. 2 from $v = -700 \text{ km s}^{-1}$ to $v = +600 \text{ km s}^{-1}$.



Extended Data Figure 6 | Spectral image and rectangular pseudo-slit slices of the UM287 data cube. **a–c,** The spectral image shown at left is in the same format as Fig. 1, and the narrow-band image that is obtained using a $-600 \text{ km s}^{-1} < v < 600 \text{ km s}^{-1}$ velocity cut is shown on the right with the corresponding slit location from which the spectral image is obtained. The vertical slits are 2.5 arcsec wide, and are positioned at 0 arcsec, 3.75 arcsec, and 5.0 arcsec in the positive x direction (northeast) with respect to the 0 reference

position. The average QSO A spectrum has been subtracted in the bright regions. In each spectral image, strong emission appears with a large, quasi-linear velocity shear centred on the QSO velocity. Also, the narrow-band image in this velocity range illustrates that, as the emission approaches the QSO, there is an offset to the northeast that is not consistent with a direct entry into the QSO. **d–f,** The same plots with the minimum intensity range set to a negative value to show negative residuals from the QSO subtraction.

High-precision comparison of the antiproton-to-proton charge-to-mass ratio

S. Ulmer¹, C. Smorra^{1,2}, A. Mooser¹, K. Franke^{1,3}, H. Nagahama^{1,4}, G. Schneider^{1,5}, T. Higuchi^{1,4}, S. Van Gorp⁶, K. Blaum³, Y. Matsuda⁴, W. Quint⁷, J. Walz^{5,8} & Y. Yamazaki⁶

Invariance under the charge, parity, time-reversal (CPT) transformation¹ is one of the fundamental symmetries of the standard model of particle physics. This CPT invariance implies that the fundamental properties of antiparticles and their matter-conjugates are identical, apart from signs. There is a deep link between CPT invariance and Lorentz symmetry—that is, the laws of nature seem to be invariant under the symmetry transformation of spacetime—although it is model dependent². A number of high-precision CPT and Lorentz invariance tests—using a co-magnetometer, a torsion pendulum and a maser, among others—have been performed³, but only a few direct high-precision CPT tests that compare the fundamental properties of matter and antimatter are available^{4–8}. Here we report high-precision cyclotron frequency comparisons of a single antiproton and a negatively charged hydrogen ion (H^-) carried out in a Penning trap system. From 13,000 frequency measurements we compare the charge-to-mass ratio for the antiproton $(q/m)_{\bar{p}}$ to that for the proton $(q/m)_p$ and obtain $(q/m)_{\bar{p}}/(q/m)_p - 1 = 1(69) \times 10^{-12}$. The measurements were performed at cyclotron frequencies of 29.6 megahertz, so our result shows that the CPT theorem holds at the atto-electronvolt scale. Our precision of 69 parts per trillion exceeds the energy resolution of previous antiproton-to-proton mass comparisons^{7,9} as well as the respective figure of merit of the standard model extension¹⁰ by a factor of four. In addition, we give a limit on sidereal variations in the measured ratio of <720 parts per trillion. By following the arguments of ref. 11, our result can be interpreted as a stringent test of the weak equivalence principle of general relativity using baryonic antimatter, and it sets a new limit on the gravitational anomaly parameter of $|\alpha_g - 1| < 8.7 \times 10^{-7}$.

The standard model is the theory that describes particles and their fundamental interactions, although without taking into account gravitation. However, this model is known to be incomplete, which has inspired searches for physics beyond the standard model, such as tests of CPT invariance that compare the fundamental properties of matter-to-antimatter equivalents at the lowest energies and with the greatest precision^{12–15}. For leptons, for example, the magnetic anomalies of electron and positron were compared with a fractional uncertainty of about 2 parts per billion⁴, and by applying similar techniques to protons and antiprotons, the resulting g -factor (a proportionality constant which links the spin of a particle to its magnetic moment) comparison reached a precision of 4.4 parts per million⁸. We are planning to improve this measurement by at least a factor of a thousand^{16,17}. In this context, we recently reported the most precise and first direct high-precision measurement of the proton magnetic moment, with a fractional precision of 3.3 parts per billion¹⁸. Complementary to these efforts, spectroscopic comparisons of hydrogen and antihydrogen are underway; recent progress has been made at CERN^{13,19}. The most precise test of CPT invariance with baryons is the comparison of the proton and antiproton charge-to-mass ratios. By measuring the cyclotron frequencies $\nu_c = (qB_0)/(2\pi m)$ of single

trapped antiprotons and H^- ions in a Penning trap with magnetic field B_0 , the TRAP collaboration⁷ achieved a fractional precision of 90 parts per trillion.

In our measurements we also compare the cyclotron frequencies of a single antiproton and an H^- ion. H^- is used as a proxy for the proton; the negative charge facilitates the experiment by eliminating the need to invert trap voltages. Our advanced Penning trap system enables adiabatic particle exchange within 15 s, which is much faster than in previous $(q/m)_{\text{H}^-/\bar{p}}$ comparisons. Our fast exchange rate allows for individual particle-to-antiparticle comparison cycles of only four minutes. This high-precision mass spectrometry method enabled us to perform about 6,500 direct frequency ratio comparisons within a total measuring time of 35 days. Moreover, our measurements have been carried out in thermal equilibrium with the detection system at 5.2(1.1) K, where systematic frequency shifts are small²⁰.

Our cryogenic Penning-trap system, which consists of a measurement trap and a reservoir trap, is shown in Fig. 1. It is mounted in the horizontal bore of a superconducting magnet at $B_0 = 1.946$ T, the axis of the magnet being oriented at 60° with respect to the Earth's rotation axis. Both traps have an inner diameter of 9 mm and are arranged in the five-electrode orthogonal and compensated design discussed in ref. 21. Transport electrodes connect the individual traps; they allow for fast adiabatic particle shuttling along the trap axis. The entire assembly is mounted in an indium-sealed cylindrical vacuum chamber with a volume of 1.2 litres. Once cooled to 4 K, ultralow pressures are reached, and thus antiproton storage lifetimes of more than a year are achieved. To measure the particle's oscillation frequencies by image current detection²² we used highly sensitive superconducting resonant detection coils²³. The measurement trap detector is operated at a resonance frequency of $\nu_{\text{res}} = 645,262$ Hz, has an inductance of 1.72 mH, and a quality factor of 11,300.

Antiprotons are delivered in bunches by the Antiproton Decelerator of CERN; the antiproton bunches also release hydrogen from our degrader structure. H^- ions are produced either by asymmetric dissociation of H_2 or by electron capture (a detailed study of the production mechanism has yet to be performed). Typically 100 to 350 cold antiprotons and about a third that number of H^- ions are prepared per ejection from the Antiproton Decelerator. From this particle cloud we extract a single antiproton and keep it in the centre of the trap, as well as an H^- ion, which is parked in the downstream park electrode. The other particles are shuttled to the reservoir trap. In case particles are lost during experiments, the measurement trap is reloaded from this reservoir.

In charge-to-mass ratio comparisons cyclotron frequencies ν_c are measured. We use the invariance theorem²⁴

$$\nu_c^2 = \nu_+^2 + \nu_z^2 + \nu_-^2 \quad (1)$$

which relates the characteristic trap frequencies ν_+ , ν_- and ν_z to ν_c . Comparisons of the antiproton-to- H^- charge-to-mass ratio are

¹RIKEN, Ulmer Initiative Research Unit, Wako, Saitama 351-0198, Japan. ²CERN, CH-1211 Geneva, Switzerland. ³Max-Planck-Institut für Kernphysik, 69117 Heidelberg, Germany. ⁴Graduate School of Arts and Sciences, University of Tokyo, Tokyo 153-8902, Japan. ⁵Institut für Physik, Johannes Gutenberg-Universität, 55099 Mainz, Germany. ⁶RIKEN, Atomic Physics Laboratory, Wako, Saitama 351-0198, Japan. ⁷GSI—Helmholtzzentrum für Schwerionenforschung, 64291 Darmstadt, Germany. ⁸Helmholtz Institut Mainz, 55099 Mainz, Germany.

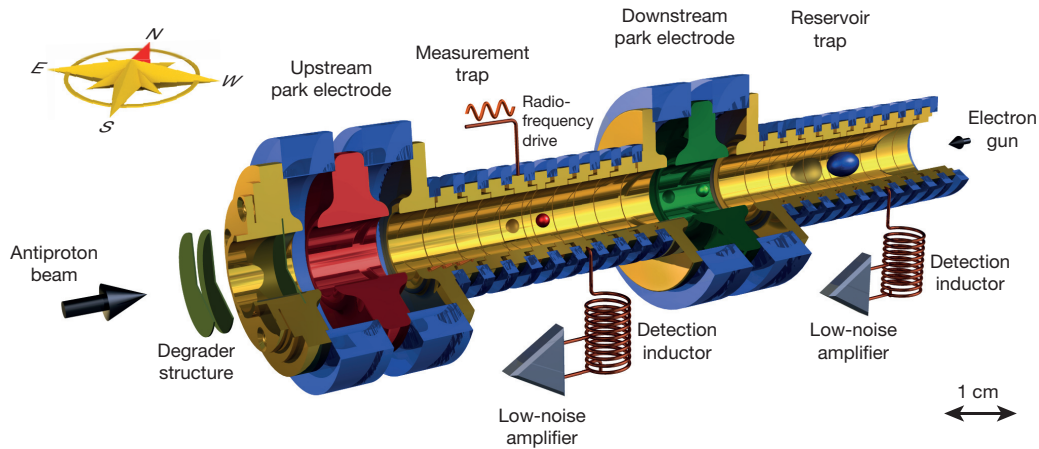


Figure 1 | Schematic of the measurement and the reservoir Penning traps. A resonant superconducting detection inductor is connected to each trap. Radio-frequency drives for particle manipulation are applied to the upstream correction electrode of the measurement trap. The upstream and downstream park electrodes are used for the particle shuttling scheme applied in the q/m

equivalent to a direct antiproton-to-proton comparison, but systematic shifts caused by polarity switching of the trapping voltages are avoided. The mass of the negative hydrogen ion is

$$m_{\text{H}^-} = m_p \left(1 + 2 \frac{m_e}{m_p} + \frac{\alpha_{\text{pol,H}^-} B_0^2}{m_p} - \frac{E_b}{m_p} - \frac{E_a}{m_p} \right) \quad (2)$$

where m_e/m_p is the electron-to-proton mass ratio²⁵, $(\alpha_{\text{pol,H}^-} B_0^2)/m_p$ is the polarizability shift⁹, E_b/m_p is the electron binding energy²⁶ and E_a/m_p is the electron affinity of hydrogen²⁷. If CPT invariance holds, the expected cyclotron frequency ratio is $R = (v_{\text{c},\bar{p}})/(v_{\text{c},\text{H}^-}) = (q/m)_{\bar{p}}/(q/m)_{\text{H}^-} = 1.001089218754(2)$, the precision being limited by the accuracy of our knowledge of the proton mass²⁸.

A detailed measurement cycle is shown in Fig. 2a. Our measurements are triggered by the antiproton injection into the Antiproton Decelerator; this avoids systematic ratio shifts induced by beats between the measurements and ambient field fluctuations caused by the Antiproton Decelerator cycle. Immediately after the injection trig-

ger, the magnetron motion of the antiproton is cooled for 10 s (ref. 18). Then the antiproton's axial frequency $v_{z,\bar{p}}$ is measured, followed by a determination of the modified cyclotron frequency at $v_{+,\bar{p}} = 29.656$ MHz, which is extracted from a sideband measurement as described in ref. 18. These two measurements take 30 s and 48 s, respectively. The particle's magnetron frequency is determined by evaluating $v_{-,\bar{p}} \approx v_{z,\bar{p}}^2/2v_{+,\bar{p}} = 7.02$ kHz, and thus $v_{\text{c},\bar{p}} \approx 29.663$ MHz is obtained. Next, we ramp the trapping potential from the configuration shown in Fig. 2b to that shown in Fig. 2c, and adjust the trapping voltage by 5 mV to tune the axial oscillation frequency of the H^- ion into resonance with the superconducting detector. Afterwards, we perform similar measurements with the hydrogen ion to determine $v_{\text{c},\text{H}^-} \approx 29.635$ MHz. Thus, a single ratio comparison takes exactly two Antiproton Decelerator cycles, corresponding to a typical measuring time of 220–240 s. To obtain from all performed measurements the final experimentally determined frequency ratio $R_{\text{exp}} = v_{\text{c},\bar{p}}/v_{\text{c},\text{H}^-} = (q/m)_{\bar{p}}/(q/m)_{\text{H}^-}$, we reject those data points measured when magnetic field changes caused by activities in the Antiproton Decelerator accelerator hall were observed. These changes are identified by an array of giant magneto resistance, Hall and flux-gate magnetic field sensors.

Subsequently the ratios are processed as follows. To remove systematic ratio shifts caused by the intrinsic magnetic-field drift $1/B_0 \times (\Delta B/\Delta t) = -5(1) \times 10^{-9}$ per hour of our superconducting magnet, we compute the antiproton cyclotron frequencies $(v_{\text{c},\bar{p},k})(t) = v_{\text{c},\bar{p},k} + (v_{\text{c},\bar{p},k+1} - v_{\text{c},\bar{p},k})/(t_{k+1} - t_k)t$, where k is the index of individual measurements. Subsequently we evaluate the upper ratio $(v_{\text{c},\bar{p},k})(t_{\text{H}^-,k})/v_{\text{c},\text{H}^-,k}$, where $t_{\text{H}^-,k}$ is the centre time of the $v_{\text{c},\text{H}^-,k}$ determination. In addition, the reciprocal ratios $v_{\text{c},\bar{p},k}/(v_{\text{c},\text{H}^-,k})(t_{\bar{p},k})$ are evaluated in a similar way. From the processed data we extract the mean of the frequency ratio by performing a maximum-likelihood fit of a Gaussian distribution to the upper ratio as well as to the reciprocal results and calculating the average of both.

To estimate the uncertainty of the mean we evaluate the correlation matrix of the extracted ratios and calculate the standard error of the cross-correlated data. This avoids underestimation of the error caused by frequencies which are, owing to the linear averaging approach, used in multiple ratios. All measured upper ratios are shown in Fig. 3a and b. The full time sequence is shown in Fig. 3a, and the results projected to a histogram are shown in Fig. 3b. Breaks in the time sequence are due to maintenance of the apparatus or systematic measurements. In Fig. 3c and d, the power spectrum density and the Allan deviation of all measured data points are shown. The mean of the power spectrum density is constant, while a linear fit to the double-log plot of the Allan

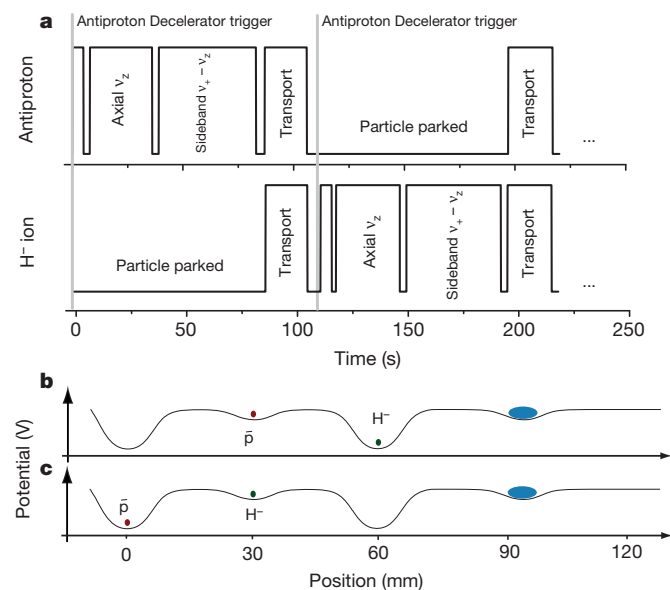


Figure 2 | Illustration of the measurement procedure. a, Detailed illustration of the measurement cycle. b, Potential configuration for antiproton cyclotron frequency measurement. The blue ellipse represents the particle reservoir. c, Potential configuration for H^- cyclotron frequency measurement.

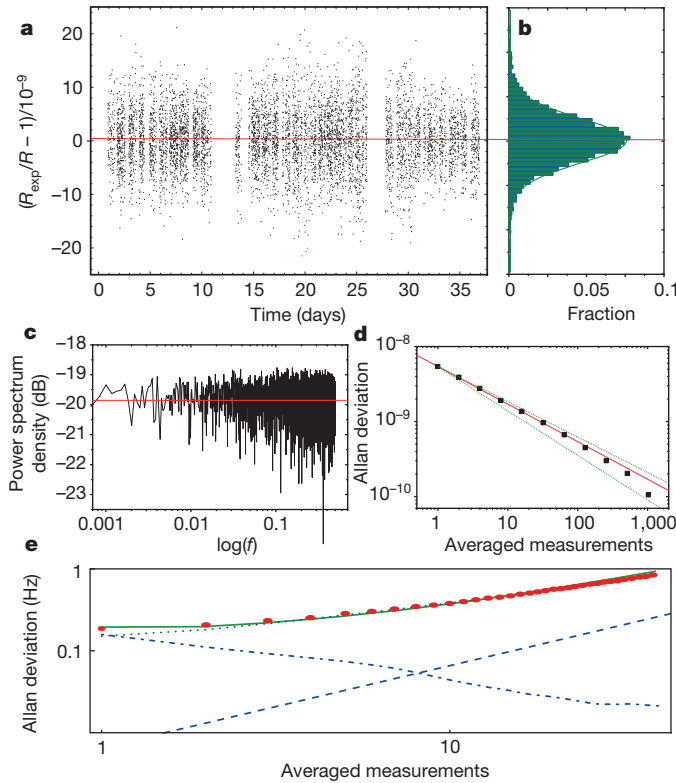


Figure 3 | Results and data analysis. **a**, All measured antiproton-to- H^- cyclotron frequency ratios as a function of time. Within 35 days, 6,521 frequency ratios were measured. **b**, Fraction, fraction of total counts (number measured in one bin normalized to the number of all measurements). **c**, Power spectrum density of the ratios as a function of reciprocal measurements. **d**, Allan deviation of the measured ratios including a fit to the data with slope $-0.501(2)$, which confirms that the ratio fluctuations follow a Gaussian white noise distribution. **e**, Allan deviation of cyclotron frequency measurements. Details are discussed in the text.

deviation gives a slope of $\alpha = -0.501(2)$. This confirms the Gaussian white-noise nature of the ratio fluctuations and justifies our data analysis.

For further verification we evaluate the Allan deviation of our cyclotron frequency measurements, which is shown in Fig. 3e as the red data points. A fit of a linear drift of $-5(1)$ parts per billion per hour (dashed line in Fig. 3e), a white-noise contribution with a root-mean-square width of $160(15)$ mHz per cycle (dash-dotted line in Fig. 3e), and a random walk generating a Gaussian distribution of $220(20)$ mHz per cycle (dotted line in Fig. 3e) reproduce the experimental results. From this model we simulate data, add random offsets Δ_i , and blind-analyse the simulation data. Within error bars the offsets Δ_i are reproduced, which independently justifies our data analysis. The antiproton-to- H^- mass ratio extracted from this data evaluation is

$$R_{\text{exp}} = 1.001089218872(64) \quad (3)$$

To further check both the data analysis and the experiment we also evaluate the cyclotron frequency ratios for \bar{p} -to- \bar{p} and H^- -to- H^- instead of \bar{p} -to- H^- , but measured within subsequent cycles. For these direct comparisons of identical particles

$$R_{\text{exp,id}} - 1 = -3(79) \times 10^{-12} \quad (4)$$

is obtained, which is consistent with 1. The increased statistical error in the second case is caused by the random walk in the magnetic field, which leads to slightly higher ratio fluctuations owing to the doubled time interval between subsequent measurements of identical particles.

Several systematic corrections enter into the measured antiproton-to- H^- charge-to-mass ratio R_{exp} . The dominant systematic shift is

related to particle exchange and the 5-mV detuning of the trapping voltage, which is required to tune the axial oscillation frequencies of both particles to the centre of the axial detector. Slightly different contact and offset potentials as well as machining imperfections are present at each individual trap electrode. Thus, the change of the ring voltage causes a relative shift of the antiproton-to- H^- equilibrium position. In the presence of a magnetic gradient term of $B_1 = 7.58(42) \text{ mT m}^{-1}$ (where B_1 is the strength of the magnetic gradient) the cyclotron frequencies $\nu_{c,\bar{p}}$ and ν_{c,H^-} are hence measured at slightly different magnetic fields, leading to a systematic ratio shift. In our measurement, the adjustment of the trapping voltage shifts the H^- ion towards lower magnetic fields, and R_{exp} has to be corrected by $-114(26)$ parts per trillion. A detailed discussion of this dominant systematic shift is provided in the Supplementary Information.

The systematic uncertainty arises from the uncertainty in the determination of the offset voltages and the magnetic gradient B_1 . In addition to the particle position, the magnitude of the octupolar correction C_4 of the trapping potential²⁰ also changes when the trap voltage is adjusted. The resulting ratio correction is at $-3(1)$ parts per trillion, the error being due to uncertainties extracted from potential theory and the determination of the axial mode temperature T_z of the particle. Eventually, the stability of our rubidium frequency reference contributes a systematic scatter of 3 parts per trillion per ratio comparison. In summary, the ratio has to be corrected by $-117(26)$ parts per trillion, leading to our final result

$$R_{\text{exp,c}} = 1.001089218755(64)(26) \quad (5)$$

which corresponds to an antiproton-to-proton mass ratio of

$$\frac{(q/m)_{\bar{p}}}{(q/m)_p} - 1 = 1(64)(26) \times 10^{-12} \quad (6)$$

and is in agreement with CPT conservation.

In the framework of the standard model extension developed by ref. 29 a figure of merit $r^{H^-} = (1 - R_{\text{exp,c}})h\nu_{c,H^-}/m_{H^-}c^2$ is derived, which characterizes the sensitivity of our measurement with respect to the CPT-violating terms added to the standard-model Lagrange density. Our result sets a new limit of $r^{H^-} < 9 \times 10^{-27}$. This exceeds the previous limit by a factor of four, and probes the standard model at an energy scale of 8 atto-electronvolts. In terms of energy sensitivity our result is the most stringent test of CPT invariance with baryonic antimatter performed so far.

In addition to the above discussion, our high data accumulation rate enables the search for sidereal variations, which might be mediated by cosmological background fields. To this end, the data set is processed using lock-in filter and Allan-deviation-based data analysis. A diurnal variation of the results would appear as a peak in the lock-in spectrum at a period of the sidereal day, which is 86,164.1 s. We do not observe such an indication and from our data analysis we conclude that the amplitude of any diurnal variation in R_{exp} is < 0.72 parts per billion at the 0.95 confidence level.

By following the arguments of ref. 11 our measurements can be interpreted as a test of weak equivalence. If the proton-to-antiproton charge-to-mass ratios are identical in the absence of a gravitational field, then the particle and antiparticle cyclotron clocks have the same frequencies. However, if matter respects weak equivalence while antimatter experiences an anomalous coupling to the gravitational field, the cyclotron frequencies of the two particles will experience different gravitational redshifts when moved to the surface of the Earth. We follow ref. 11, in which a possible gravitational anomaly acting on antimatter is expressed by a parameter α_g , which modifies the effective newtonian gravitational potential U to give $\alpha_g U$. This contributes a possible difference in the measured cyclotron frequencies of $(\nu_{c,p} - \nu_{c,\bar{p}})/(\nu_{c,p}) = -3(\alpha_g - 1)U/c^2$. Thus, our measurement sets a new upper limit of $|\alpha_g - 1| < 8.7 \times 10^{-7}$.

Received 26 May; accepted 17 June 2015.

1. Lüders, G. Proof of the TCP theorem. *Ann. Phys.* **2**, 1–15 (1957).
2. Greenberg, O. W. CPT violation implies Lorentz violation. *Phys. Rev. Lett.* **89**, 231602 (2002).
3. Kostelecký, V. A. & Russell, N. Data-tables on Lorentz and CPT Violation. *Rev. Mod. Phys.* **83**, 11–31 (2011).
4. Van Dyck, R., Jr, Schwinberg, P. & Dehmelt, H. New high-precision comparison of electron and positron g factors. *Phys. Rev. Lett.* **59**, 26–29 (1987).
5. Schwingerheuer, B. *et al.* CPT tests in the neutral Kaon system. *Phys. Rev. Lett.* **74**, 4376–4379 (1995).
6. Miller, J. P., De Rafael, E. & Roberts, B. L. Muon ($g-2$): experiment and theory. *Rep. Prog. Phys.* **70**, 795 (2007).
7. Gabrielse, G. *et al.* Precision mass spectroscopy of the antiproton and proton using simultaneously trapped particles. *Phys. Rev. Lett.* **82**, 3198–3201 (1999).
8. DiSciaccia, J. *et al.* One-particle measurement of the antiproton magnetic moment. *Phys. Rev. Lett.* **110**, 130801 (2013).
9. Thompson, J. K., Rainville, S. & Pritchard, D. E. Cyclotron frequency shifts arising from polarization forces. *Nature* **430**, 58–61 (2004).
10. Colladay, D. & Kostelecký, V. A. CPT violation and the standard model. *Phys. Rev. D* **55**, 6760–6774 (1997).
11. Hughes, R. J. & Holzscneider, M. H. Constraints on the gravitational properties of antiprotons and positrons from cyclotron-frequency measurements. *Phys. Rev. Lett.* **66**, 854–857 (1991).
12. Hori, M. *et al.* Two-photon spectroscopy of antiprotonic helium and the electron-to-antiproton mass ratio. *Nature* **475**, 484–488 (2011).
13. Kuroda, N. *et al.* A source of antihydrogen atoms for in flight hyperfine spectroscopy. *Nature Commun.* **5**, 3089–3092 (2014).
14. Andresen, G. B. *et al.* Confinement of antihydrogen for 1,000 seconds. *Nature Phys.* **7**, 558–564 (2011).
15. Gabrielse, G. *et al.* Background-free observation of cold antihydrogen and a field-ionization analysis of its states. *Phys. Rev. Lett.* **89**, 213401 (2002).
16. Ulmer, S. *et al.* CERN Document Server Report SPSC-TDR-002, <http://cds.cern.ch/record/1503514?ln5de> (CERN, 2013).
17. Smorra, C. *et al.* Towards a high-precision measurement of the antiproton magnetic moment. *Hyperfine Interact.* **228**, 31–36 (2014).
18. Mooser, A. *et al.* High-precision measurement of the magnetic moment of the proton. *Nature* **509**, 596–599 (2014).
19. Amole, C. *et al.* Resonant quantum transitions in trapped antihydrogen atoms. *Nature* **483**, 439–443 (2012).
20. Brown, L. S. & Gabrielse, G. Geonium theory: physics of a single electron or ion in a Penning trap. *Rev. Mod. Phys.* **58**, 233–311 (1986).
21. Gabrielse, G., Haarsma, L. & Rolston, L. S. Open-endcap Penning traps for high precision experiments. *Int. J. Mass Spectrom.* **88**, 319–332 (1989).
22. Wineland, D. J. & Dehmelt, H. G. The stored ion calorimeter. *J. Appl. Phys.* **46**, 919–922 (1975).
23. Ulmer, S. *et al.* The quality factor of a superconducting resonator in a magnetic field. *Rev. Sci. Instrum.* **80**, 123302 (2009).
24. Brown, L. S. & Gabrielse, G. Precision spectroscopy of a charged particle in an imperfect Penning trap. *Phys. Rev. A* **25**, 2423–2425 (1982).
25. Sturm, S. *et al.* High-precision measurement of the atomic mass of the electron. *Nature* **506**, 467–470 (2014).
26. Parthey, C. G. *et al.* Improved measurement of the hydrogen 1S–2S transition frequency. *Phys. Rev. Lett.* **107**, 203001 (2011).
27. Bhatia, A. K. & Drachmann, R. J. Polarizability of helium and the negative hydrogen ion. *J. Phys. At. Mol. Opt. Phys.* **27**, 1299–1305 (1994).
28. Van Dyck, R. S. Jr, Farnham, D. L., Zafonte, S. L. & Schwinberg, P. B. High precision Penning trap mass spectroscopy and a new measurement of the proton's atomic mass. *AIP Conf. Proc.* **457**, 101–110 (1999).
29. Bluhm, R., Kostelecký, V. A. & Russell, N. CPT and Lorentz tests in Penning traps. *Phys. Rev. D* **57**, 3932–3943 (1998).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge support by the Antiproton Decelerator group and all other CERN groups which provide support to Antiproton Decelerator experiments. We thank L. Bojtar and F. Butin for their invaluable support. We acknowledge financial support by the RIKEN Initiative Research Unit Program, RIKEN President Funding, RIKEN Pioneering Project Funding, RIKEN FPR Funding, the RIKEN JRA Program, the Grant-in-Aid for Specially Promoted Research (grant number 24000008) of MEXT, the Max-Planck Society, the IMPRS-PTFS, the EU (ERC advanced grant number 290870-MEFUCO), the BMBF, the Helmholtz-Gemeinschaft, and the CERN Fellowship programme.

Author Contributions S.U., C.S. and G.S. designed the apparatus which was then developed and commissioned by S.U., C.S., G.S., K.F., H.N., T.H. and A.M.; S.U., A.M. and C.S. developed the software system and experiment control. S.U., A.M., C.S., H.N., T.H., K.F. and G.S. participated in the 2014 antiproton run and contributed to the data collection. C.S., A.M., and S.U. discussed and analysed the data. S.U. and A.M. performed the systematic studies. The manuscript was written by S.U., A.M., K.B. and J.W. and then discussed and approved by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.U. (stefan.ulmer@cern.ch).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

Rejuvenation of metallic glasses by non-affine thermal strain

S. V. Ketov^{1*}, Y. H. Sun^{2*}, S. Nachum², Z. Lu³, A. Checchi^{2,4}, A. R. Beraldin^{2,4}, H. Y. Bai³, W. H. Wang³, D. V. Louzguine-Luzgin¹, M. A. Carpenter⁵ & A. L. Greer^{1,2}

When a spatially uniform temperature change is imposed on a solid with more than one phase, or on a polycrystal of a single, non-cubic phase (showing anisotropic expansion-contraction), the resulting thermal strain is inhomogeneous (non-affine). Thermal cycling induces internal stresses, leading to structural and property changes that are usually deleterious. Glasses are the solids that form on cooling a liquid if crystallization is avoided—they might be considered the ultimate, uniform solids, without the microstructural features and defects associated with polycrystals. Here we explore the effects of cryogenic thermal cycling on glasses, specifically metallic glasses. We show that, contrary to the null effect expected from uniformity, thermal cycling induces rejuvenation, reaching less relaxed states of higher energy. We interpret these findings in the context that the dynamics in liquids become heterogeneous on cooling towards the glass transition¹, and that there may be consequent heterogeneities in the resulting glasses. For example, the vibrational dynamics of glassy silica at long wavelengths are those of an elastic continuum, but at wavelengths less than approximately three nanometres the vibrational dynamics are similar to those of a polycrystal with anisotropic grains². Thermal cycling of metallic glasses is easily applied, and gives improvements in compressive plasticity. The fact that such effects can be achieved is attributed to intrinsic non-uniformity of the glass structure, giving a non-uniform coefficient of thermal expansion. While metallic glasses may be particularly suitable for thermal cycling, the non-affine nature of strains in glasses in general deserves further study, whether they are induced by applied stresses or by temperature change.

In glasses there are no equivalent atoms, and elastic deformation must be non-affine. For metallic glasses this has been explored through atomistic simulations^{3,4} and diffraction-based measurements^{5,6} of the pair distribution function (see, for example, the review in ref. 7). Compared to their crystalline counterparts, metallic glasses have lower elastic moduli, associated with local atomic rearrangements occurring even well within the nominally elastic regime⁸. While these rearrangements (shears) contribute to apparently elastic strain, they are not individually reversible⁷, and there is net structural change. In atomistic simulations, shear events are localized in ~ 1 nm soft spots with low-frequency vibrational modes⁴, and can be associated with the range of density and stiffness of nearest-neighbour clusters (that is, the shell around a central atom)^{3,4}.

Measurements of metallic glass properties do not have such fine spatial resolution, but non-uniformity is still clearly detected. Strongly and weakly bonded regions have been postulated to account for ~ 5 nm zones of accelerated crystallization⁹. The initial yield load (in nano-indentation) shows a dispersion of values associated with structural heterogeneity¹⁰. Mapping of heterogeneity by atomic-force microscopy, over areas much larger than in simulation, shows variations of

energy dissipation¹¹ and of elastic modulus ($\sim 30\%$)¹² with correlation lengths of 2.5–20 nm.

In metallic glasses under elastic loading, non-affine strains lead to local structural change; this led us to question whether thermal strains show analogous effects. In elastic deformation, the effects of non-affine strain and modulus variation are strongest for shear. For thermal expansion/contraction of a glass, however, the macroscopic strain is hydrostatic. In simulations of amorphous iron, the local shear modulus G shows a relative standard deviation of 27%; the local bulk modulus B shows a smaller, but still substantial, value of 18%¹³. This variation is likely to be greater for the range of bonding found in a multicomponent system¹⁴. As the coefficient of thermal expansion (CTE) is—other factors being constant—inversely proportional to B (ref. 15), we expect that it also shows significant variation. Noting that neighbouring regions of different CTE must satisfy elastic compatibility, atomic-level shears must develop on temperature change, and given the existence of soft spots, local structural change may occur. On annealing, metallic glasses undergo thermally activated structural relaxation to states of higher density and lower enthalpy. To avoid such effects, thermal cycling in the present work is only from near room temperature (293–343 K) to lower temperatures, always far below the glass-transition temperature T_g near which thermal relaxation occurs (Fig. 1)¹⁶.

As a metallic glass is heated towards T_g , there is exothermic structural relaxation (Fig. 2a). The relaxation spectrum (heat release rate as a function of temperature) can be detected by differential scanning calorimetry (DSC). While the shape of the spectrum may be of interest¹⁷, we focus here only on the overall heat of relaxation (ΔH_{rel} , shaded areas in Fig. 2b). Changes in ΔH_{rel} have been used to characterize the effects of thermal and mechanical treatments on metallic glasses; for example, heavy plastic deformation (shot-peening¹⁷) can induce ageing (that is, relaxation to a lower-energy state) or rejuvenation. For lanthanum (La)-based metallic glasses, cycling between room temperature (293 K) and liquid-nitrogen temperature (77 K) leads to increases in ΔH_{rel} , that is, rejuvenation (Fig. 2b). For melt-spun ribbons, ΔH_{rel} peaks at ~ 10 cycles; for bulk glass, ΔH_{rel} is smaller, peaking at ~ 15 cycles; in each case, the maximum increase in ΔH_{rel} is $\sim 50\%$ relative to the as-cast sample.

We next addressed whether such changes are due to cycling, or to the time spent at 77 K. Ribbons were held at 77 K for 4 h, much longer than the sum of the times spent at 77 K in cycling treatments, but this ‘anneal’ did not give a detectable change in ΔH_{rel} (see Methods and Extended Data Fig. 1).

Because surfaces respond more rapidly than interiors, sample temperature is not spatially uniform during thermal cycling; surfaces must be in tension during cooling and in compression during heating. In nano-indentation, stress cycling can harden metallic glasses¹⁸, raising the question of whether the ΔH_{rel} changes (Fig. 2b) are due to the

¹WPI Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan. ²University of Cambridge, Department of Materials Science & Metallurgy, 27 Charles Babbage Road, Cambridge CB3 0FS, UK. ³Institute of Physics, Chinese Academy of Sciences, Beijing 100080, China. ⁴Department of Management and Engineering, University of Padua, 3 Stradella San Nicola, Vicenza 36100, Italy. ⁵University of Cambridge, Department of Earth Sciences, Cambridge CB2 3EQ, UK.

*These authors contributed equally to this work.

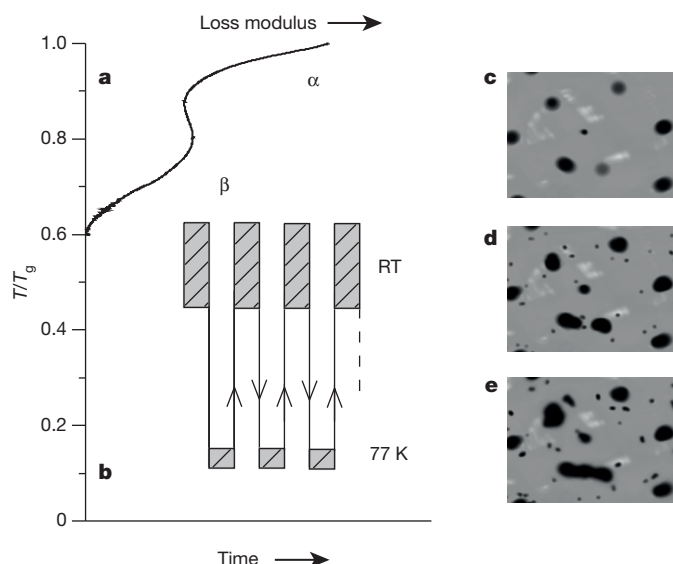


Figure 1 | Thermal cycling of metallic glasses. **a**, At high fractions of the glass-transition temperature, T_g , to as low as $0.6 T_g$, metallic glasses can undergo α and β relaxations, giving peaks in loss modulus as a function of T/T_g as shown by the curve¹⁶. **b**, The thermal cycling in the present work explores lower temperatures. Samples are cycled from near room temperature, 293–343 K (RT), to liquid-nitrogen temperature (77 K). The ranges of T/T_g reflect both range of temperature and the range of T_g for the metallic glasses, from 474 K for $\text{La}_{55}\text{Ni}_{10}\text{Al}_{35}$ (at.%) up to 700 K for $\text{Cu}_{46}\text{Zr}_{46}\text{Al}_7\text{Gd}_1$. **c–e**, Schematic depictions of the degree of heterogeneity in a metallic glass in the as-cast state (**c**), and after increasing numbers of thermal cycles (**d**, **e**). The population and intensity of soft spots (dark), with lower elastic stiffness and higher CTE, increases with cycling. The scale of these heterogeneities is not given by the present results (so no scale bars are given), but mapping of elastic properties^{11,12} suggests characteristic lengths of <10 nm.

temperature cycling itself, or to the associated stresses. As-cast rods of $\text{La}_{55}\text{Ni}_{10}\text{Al}_{35}$ bulk metallic glass (BMG) were subjected to room temperature–77 K cycles at the same time as thin discs cut from the rods. As detailed in Methods, the changes in ΔH_{rel} are independent of sample size, showing that stresses generated during cycling have a negligible effect (Extended Data Fig. 2). Thermal calculations (Extended Data Table 1) suggest that the stresses themselves are negligible, consistent with this conclusion.

Many aspects of thermal cycling could be tailored: upper and lower temperatures, holding times at these temperatures, rates of cooling and heating, number of cycles. We report preliminary studies in Methods; our focus here, however, is on property changes induced by thermal cycling.

In instrumented indentation (nano-indentation) of metallic glasses^{10,18}, initial yielding is indicated by a sharp ‘pop-in’ (increase in indentation depth h) on the load– h curve (Extended Data Fig. 3), corresponding to shear-banding onset. Cumulative distributions of initial yield pressure P_y (Fig. 3a) for $\text{La}_{55}\text{Ni}_{20}\text{Al}_{25}$ glass ribbon in three states permit the effects of a longer hold at 77 K to be compared with cycling. The median value of P_y , 2.98 GPa in the as-cast sample, decreases by 3% after 10-min hold at 77 K, and by a further 17% after ten room temperature–77 K (1 min) cycles. After the cycling treatments the relative width (1st to 9th decile) of the P_y distribution ($\pm 15\%$) is greater than in the as-cast sample ($\pm 7\%$), suggesting that cycling induces greater heterogeneity.

The corresponding distributions of hardness H (Fig. 3b), determined from h at full load, show similar trends. The median H of the as-cast sample, 2.65 GPa, is 1% lower after 10 min at 77 K, and a further 4% lower after ten room temperature–77 K cycles. For a given sample, the H distribution is narrower than that for P_y ; for example, after cycling, the relative width is only $\pm 3\%$ (one-fifth that of P_y). Evidently H , reflecting conditions for continuance of flow, is much

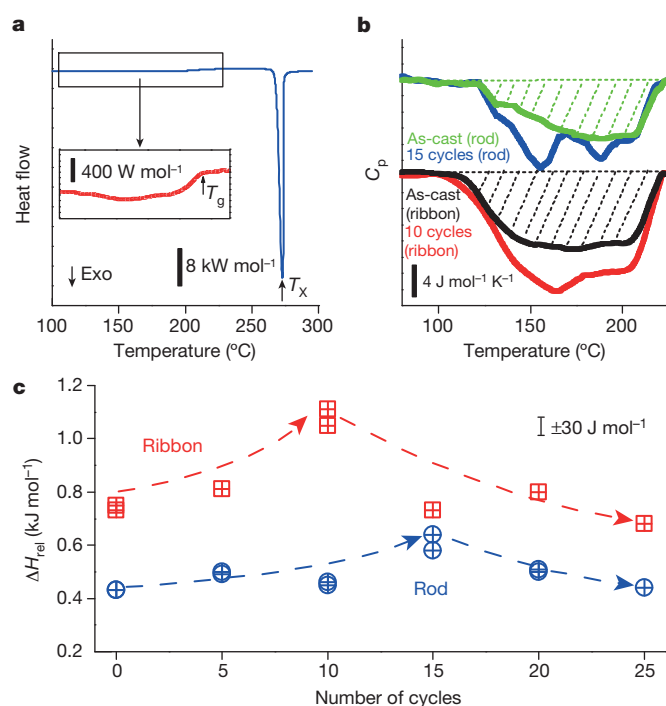


Figure 2 | Differential scanning calorimetry (DSC) of melt-spun ribbons of $\text{La}_{55}\text{Ni}_{20}\text{Al}_{25}$ and bulk rods of $\text{La}_{55}\text{Ni}_{10}\text{Al}_{35}$ metallic glasses. **a**, DSC traces on heating show an exotherm before the glass transition at T_g (here for ribbon). **b**, By subtraction of successive traces of specific heat C_p (see Methods), the heat of relaxation, ΔH_{rel} , associated with the exotherm can be determined (shaded areas) and is increased after room temperature–77 K thermal cycles. In both **a** and **b**, vertical bars give units and scales for the y axes. **c**, For ribbons, ΔH_{rel} peaks at ~ 10 room temperature–77 K cycles; for discs (250–500 μm thick) cut from the bulk rod, ΔH_{rel} is lower and peaks at ~ 15 cycles; the data points are for individual measurements and the changes lie outside the error range of ± 30 J mol^{−1} (standard deviation estimated as in Methods).

less sensitive to local heterogeneities than is P_y . The Young modulus, determined on unloading, shows similar behaviour to H (Extended Data Fig. 4).

The distribution curves of P_y and H confirm changes related to the number of cycles rather than the holding time at 77 K. The marked reduction in P_y induced by cycling is accompanied by a reduction in initial yield increments (Δh , Extended Data Fig. 4a).

Metallic glasses show exceptionally high yield stress and yield strain¹⁹, and can be very tough²⁰, but there is a clear need to improve their plasticity. Annealing reduces their plasticity, often inducing brittleness. Figures 2 and 3 show property changes opposite to those expected from annealing, suggesting that thermal cycling gives rejuvenation. The effects on macroscopic plasticity are explored for two BMG compositions tested in uniaxial compression (Fig. 4). The plastic strain of 1.5-mm-diameter samples of $\text{Zr}_{62}\text{Cu}_{24}\text{Fe}_5\text{Al}_9$ is improved (from 4.9% to 7.6%) by successive thermal cycles (Fig. 4a). Improvements in plasticity, smaller in absolute terms (for reasons given in Methods), but larger relatively, are seen in larger-diameter samples of the same glass (Fig. 4b), which also show a reduction in microhardness (by 4%, inset), similar to that for the median hardness of the La-based glass in Fig. 3b. $\text{Cu}_{46}\text{Zr}_{46}\text{Al}_7\text{Gd}_1$ BMG (1.5 × 1.5 × 3.0 mm³ cuboid) shows an increase in plastic strain from 1.4% in the as-cast state to 5.1% after ten room temperature–77 K cycles (Fig. 4c). A sample fully relaxed by annealing is brittle and is not successfully rejuvenated by cycling, but a strong rejuvenation effect is possible after partial thermal relaxation, when thermal cycling gives a plastic strain similarly improved to that of as-cast samples.

Well below T_g , plastic flow in metallic glasses is localized into thin shear bands. Macroscopic plasticity is limited by premature failure on

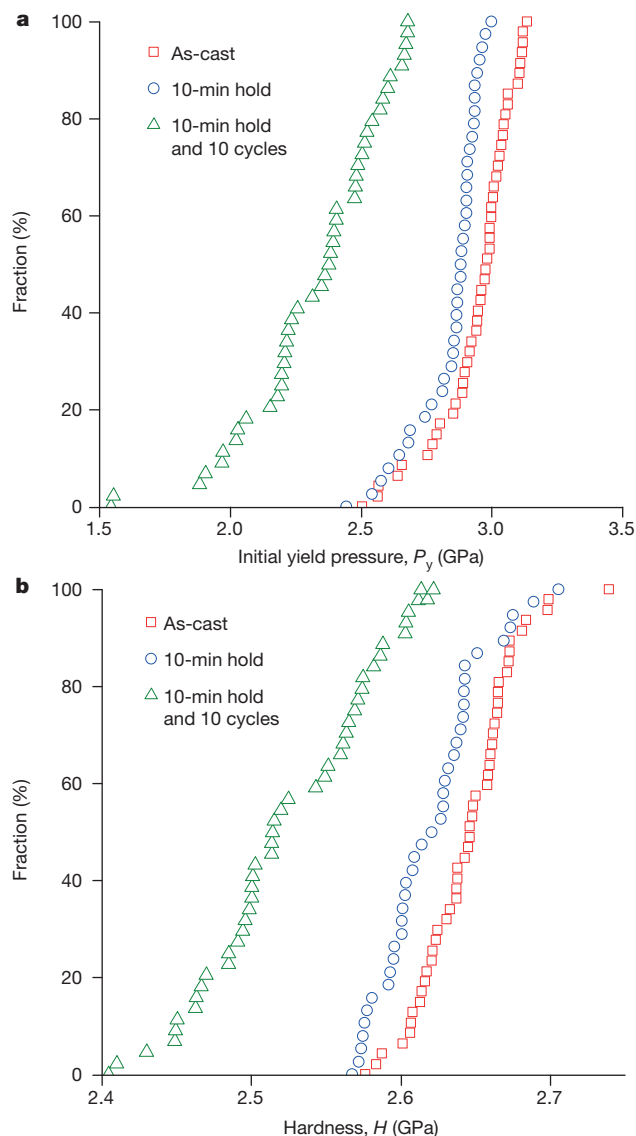


Figure 3 | Cumulative distributions of (a) initial yield pressure P_y and (b) hardness H . $\text{La}_{55}\text{Ni}_{20}\text{Al}_{25}$ metallic glass ribbon is tested in instrumented indentation, to a maximum load of 40 mN, in three states: as-cast, after a 10-min hold at 77 K, and after a further ten room temperature–77 K cycles each with 1 min hold. The cycling reduces P_y and H , and widens the distribution of their values. The widening is taken to represent increasing heterogeneity of the glass structure (Fig. 1c–e).

one or a few dominant shear bands. The plasticity can be improved by proliferation of shear bands, making the overall flow more uniform²¹. The shear-band spacing of $\sim 10\ \mu\text{m}$ in the as-cast sample (Fig. 4d) is reduced to $\sim 2.5\ \mu\text{m}$ in the thermally cycled sample, suggesting greater ease of shear-band initiation, consistent with reductions in P_y (Fig. 3a) and in initial yield displacements (Extended Data Fig. 4a). Resonant ultrasound spectroscopy (RUS) was used to measure the elastic moduli of $\text{Cu}_{46}\text{Zr}_{46}\text{Al}_7\text{Gd}_1$ BMG, which show no detectable change on thermal cycling (Extended Data Fig. 5).

Across a range of metallic glass compositions, melt-spun and bulk, thermal cycling induces changes in properties. As the cycling is down to 77 K, these changes can be compared with the property improvements (hardness, wear resistance) obtained by deep cryogenic treatment (DCT) of steels²². DCT, however, works through phase changes (transformation of residual austenite, precipitation of fine carbides) on holding at low temperature. In the present case, there is no discernible phase change (X-ray diffraction confirms that the samples remain fully

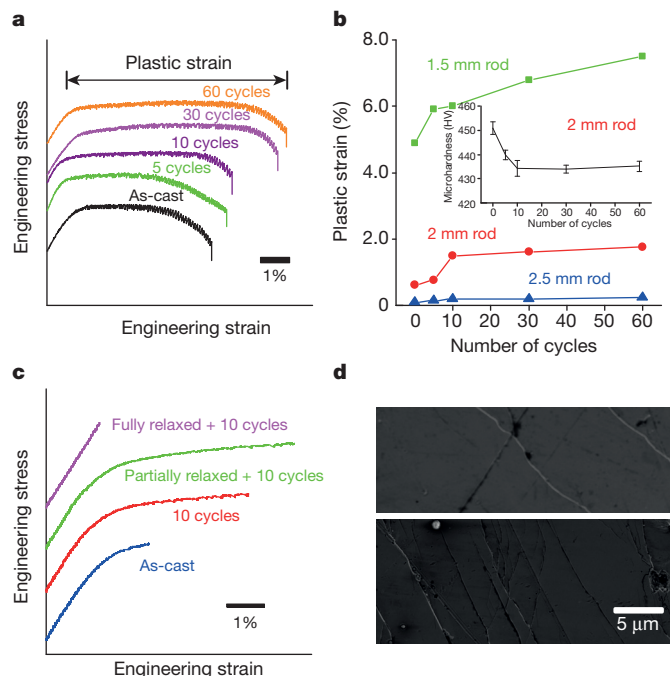


Figure 4 | Improved plasticity after thermal cycling. **a**, 1.5-mm-diameter rod samples of $\text{Zr}_{62}\text{Cu}_{24}\text{Fe}_5\text{Al}_9$ BMG (yield stress $1.68 \pm 0.05\ \text{GPa}$) show an increase in plastic strain under uniaxial compression when treated with 338 K to 77 K thermal cycles. **b**, The compressive plasticity of this glass is strongly dependent on sample dimension, but for all the diameters tested, thermal cycling leads to an increase in plastic strain, accompanied (inset) by a reduction in microhardness (in kg mm^{-2} ; each data point is the average of 20 measurements). **c**, $1.5 \times 1.5 \times 3.0\ \text{mm}^3$ cuboids of $\text{Cu}_{46}\text{Zr}_{46}\text{Al}_7\text{Gd}_1$ BMG (yield stress $1.61 \pm 0.04\ \text{GPa}$), both as-cast and annealed (1.0 h at 400°C) to partial relaxation, show an increase in plastic strain under uniaxial compression after ten room temperature–77 K cycles; for full relaxation (1.5 h at 400°C), the sample remains brittle. **d**, For the $\text{Cu}_{46}\text{Zr}_{46}\text{Al}_7\text{Gd}_1$ BMG samples, scanning electron microscopy shows that the population density of shear bands near the dominant shear band is lower in an as-cast sample (top) than in a similar sample tested after ten room temperature–77 K cycles (bottom). In **a** and **c**, horizontal bars indicate the scale and units for the x axes; quoted yield stress gives a guide to the y axes.

glassy after cycling; Extended Data Fig. 6), and the effects are more analogous to those induced in single-phase polycrystalline non-cubic metals, for example, thermal-cycling growth in uranium²³ and increased dislocation density in zirconium²⁴. The widespread use of DCT²⁵ does, however, suggest that cycling to cryogenic temperatures may be a practicable process.

The changes in ΔH_{rel} (Fig. 2c) show rejuvenation followed by ageing, opposing trends known in processes such as irradiation or plastic deformation, which introduce both damage (raising the energy of the system) and mobility (enabling relaxation to lower energy). Shot-peening increases ΔH_{rel} in an annealed BMG, but for an as-cast BMG even this intense plastic deformation reduces ΔH_{rel} (ref. 17). Thermal cycling increases ΔH_{rel} even for a rapidly quenched metallic glass. The maximum increase in ΔH_{rel} (Fig. 2c), $340\ \text{J mol}^{-1}$, is (correcting for differing DSC heating rates) $\sim 60\%$ of the increase in ΔH_{rel} in a Zr-based BMG induced by one rotation in high-pressure torsion²⁶, and reductions in hardness (Figs 3b and 4b and ref. 26) are, pro rata, similar. The increases in ΔH_{rel} (Fig. 2c) are very similar to those seen in elastostatic loading²⁷; it is remarkable that both these apparently mild processes give effects comparable with heavy plastic deformation, suggesting that non-affine strains in the nominally elastic regime are efficient in generating structural damage (disordering).

Although continued cycling leads to reversal in the ΔH_{rel} increase (Fig. 2c), the microhardness reduction appears to saturate (Fig. 4b inset) while the plasticity continues to improve (Fig. 4a, b); the onset

of yield (Fig. 3a) is facilitated more than general yielding (Fig. 3b) and the effect on high-frequency (MHz in RUS) macroscopic elastic moduli is negligible (Extended Data Fig. 5). These contrasts for different properties lead us to suggest that thermal cycling introduces heterogeneities (soft spots)⁴ that are particularly effective in initiating flow and improving plasticity, with lesser effects on the average structure of the metallic glass. That the effects (at least for ΔH_{rel} , Fig. 2c) are greater in melt-spun ribbon than in BMG of similar composition suggests that pre-existing heterogeneity (greater in a less-relaxed metallic glass) helps cycling to introduce more heterogeneity; this is expected if, as we speculate, the effect is due to non-affine thermal strains arising from the heterogeneity itself. The present results, however, do not identify any characteristic length scale for the heterogeneity.

Rejuvenation of metallic glasses, with improved plasticity, has been achieved by elastostatic loading^{27,28}, ion irradiation²⁹ and plastic deformation²¹. In comparison with these methods, thermal cycling is attractive: it is non-destructive, not involving shape change; unlike an applied elastostatic stress, it is isotropic and cannot introduce anisotropy; it can be applied to any sample (thin film, ribbon or bulk) and repeated as necessary; it changes the whole sample (not just the surface as, for example, in shot-peening or ion irradiation, or only inside shear bands); and it is controllable, can be performed *in situ*, induces no macroscopic residual stresses, involves strains far below the elastic limit, and induces no macroscopic plastic flow.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 January; accepted 8 June 2015.

- Ediger, M. D. Spatially heterogeneous dynamics in supercooled liquids. *Annu. Rev. Phys. Chem.* **51**, 99–128 (2000).
- Baldi, G. *et al.* Emergence of crystal-like atomic dynamics in glasses at the nanometer scale. *Phys. Rev. Lett.* **110**, 185503 (2013).
- Ding, J., Cheng, Y. Q. & Ma, E. Correlating local structure with inhomogeneous elastic deformation in a metallic glass. *Appl. Phys. Lett.* **101**, 121917 (2012).
- Ding, J., Patinet, S., Falk, M. L., Cheng, Y. & Ma, E. Soft spots and their structural signature in a metallic glass. *Proc. Natl Acad. Sci. USA* **111**, 14052–14056 (2014).
- Poulsen, H. F., Wert, J. A., Neufeind, J., Honkimaäki, V. & Daymond, M. Measuring strain distributions in amorphous materials. *Nature Mater.* **4**, 33–36 (2005).
- Hufnagel, T. C., Ott, R. T. & Almer, J. Structural aspects of elastic deformation of a metallic glass. *Phys. Rev. B* **73**, 064204 (2006).
- Egami, T., Iwashita, T. & Dmowski, W. Mechanical properties of metallic glasses. *Metals* **3**, 77–113 (2013).
- Weaire, D., Ashby, M. F., Logan, J. & Weins, M. J. On the use of pair potentials to calculate the properties of amorphous metals. *Acta Metall.* **19**, 779–788 (1971).
- Ichitsubo, T. *et al.* Microstructure of fragile metallic glasses inferred from ultrasound-accelerated crystallization in Pd-based metallic glasses. *Phys. Rev. Lett.* **95**, 245501 (2005).
- Packard, C. E., Franke, O., Homer, E. R. & Schuh, C. A. Nanoscale strength distribution in amorphous versus crystalline metals. *J. Mater. Res.* **25**, 2251–2263 (2010).
- Liu, Y. H. *et al.* Characterization of nanoscale mechanical heterogeneity in a metallic glass by dynamic force microscopy. *Phys. Rev. Lett.* **106**, 125504 (2011).
- Wagner, H. *et al.* Local elastic properties of a metallic glass. *Nature Mater.* **10**, 439–442 (2011).
- Egami, T. & Srolovitz, D. Local structural fluctuations in amorphous and liquid metals: a simple theory of the glass transition. *J. Phys. F Met. Phys.* **12**, 2141–2163 (1982).
- Mendelev, M. I., Ott, R. T., Kramer, M. J. & Sordet, D. J. Determining strain in amorphous alloys: Uncertainties with analyzing structural changes during deformation. *J. Appl. Phys.* **105**, 023509 (2009).
- Grüneisen, E. Theorie des festen Zustandes ein-atomiger element. *Ann. Phys.* **344**, 257–306 (1912).
- Yu, H. B., Wang, W. H. & Samwer, K. The β relaxation in metallic glasses: an overview. *Mater. Today* **16**, 183–191 (2013).
- Concustell, A., Méar, F. O., Suriñach, S., Baró, M. D. & Greer, A. L. Structural relaxation and rejuvenation in a metallic glass induced by shot-peening. *Philos. Mag. Lett.* **89**, 831–840 (2009).
- Packard, C. E., Homer, E. R., Al-Aqeeli, N. & Schuh, C. A. Cyclic hardening of metallic glasses under Hertzian contacts: Experiments and STZ dynamics simulations. *Phil. Mag.* **90**, 1373–1390 (2010).
- Ashby, M. F. & Greer, A. L. Metallic glasses as structural materials. *Scr. Mater.* **54**, 321–326 (2006).
- Demetriou, M. D. *et al.* A damage-tolerant glass. *Nature Mater.* **10**, 123–128 (2011).
- Greer, A. L., Cheng, Y. Q. & Ma, E. Shear bands in metallic glasses. *Mater. Sci. Eng. Rep.* **74**, 71–132 (2013).
- Preciado, M. & Pellizarri, M. Influence of deep cryogenic treatment on the thermal decomposition of Fe–C martensite. *J. Mater. Sci.* **49**, 8183–8191 (2014).
- Ma, B. M. *Nuclear Reactor Materials and Applications* Sect. 6.7.2., 149–151 (Van Nostrand, 1983).
- Yuan, C. *et al.* Effects of deep cryogenic treatment on the microstructure and mechanical properties of commercial pure zirconium. *J. Alloy. Comp.* **619**, 513–519 (2015).
- Kalsi, N. S., Sehgal, R. & Sharma, V. S. Cryogenic treatment of tool materials: a review. *Mater. Manuf. Process.* **25**, 1077–1100 (2010).
- Meng, F., Tsuchiya, K., Seichiro, I. & Yokoyama, Y. Reversible transition of deformation mode by structural rejuvenation and relaxation in bulk metallic glass. *Appl. Phys. Lett.* **101**, 121914 (2012).
- Lee, S. C., Lee, C. M., Yang, J. W. & Lee, J. C. Microstructural evolution of an elastostatically compressed amorphous alloy and its influence on the mechanical properties. *Scr. Mater.* **58**, 591–594 (2008).
- Ke, H. B., Wen, P., Peng, H. L., Wang, W. H. & Greer, A. L. Homogeneous deformation of metallic glass at room temperature reveals large dilatation. *Scr. Mater.* **64**, 966–969 (2011).
- Magagnosc, D. J. *et al.* Effect of ion irradiation on tensile ductility, strength and fictive temperature in metallic glass nanowires. *Acta Mater.* **74**, 165–182 (2014).

Acknowledgements This research was supported by the World Premier International Research Center Initiative (WPI), MEXT, Japan, by NSF China and MOST 973 China, and by the Engineering and the Engineering and Physical Sciences Research Council, UK (Materials World Network project). Y.H.S. acknowledges support from a China Scholarship Council (CSC) scholarship. M. L. Falk, T. C. Hufnagel, E. Ma, D. B. Miracle and J. Orava are thanked for discussions. All data accompanying this publication are directly available within the publication and the accompanying Extended Data figures and tables.

Author Contributions Preparation of metallic glasses was by S.V.K. and Z.L. The thermal-cycling treatments were performed by Y.H.S., S.V.K., A.R.B. and A.C.; calorimetry by Y.H.S., S.V.K., Z.L., A.C. and A.R.B.; nano-indentation by S.N. and A.R.B.; compression and microhardness tests by S.V.K., Y.H.S., A.R.B. and A.C.; X-ray diffraction by S.V.K. and Y.H.S.; and resonant ultrasound spectroscopy by Y.H.S., M.A.C., A.R.B. and A.C. Direction of the work was by D.V.L.-L. (Sendai), H.Y.B. and W.H.W. (Beijing), and A.L.G. (Cambridge). A.L.G. led the project and the writing of the paper. All authors contributed to interpretation and presentation of the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.L.G. (alg13@cam.ac.uk).

METHODS

Metallic glasses in this study. Sample preparation and property measurements took place in three laboratories. Kinetic phenomena in glasses are best compared on a homologous temperature scale, using the glass-transition temperature T_g as the normalizing parameter. The selected glasses allow studies over a range of T/T_g ; for example, room temperature (293 K) covers the range of $T/T_g = 0.42$ – 0.62 for these glasses. (The values of T_g were determined using differential scanning calorimetry at standard heating rates of 20 and 40 K min⁻¹.) The four selected glasses are: (1) Cu₄₆Zr₄₆Al₇Gd₁ rod, 3 mm diameter ($T_g = 700$ K (ref. 30)); (2) La₅₅Ni₂₀Al₂₅ (at.%) melt-spun ribbon 40 µm thick ($T_g = 475$ K); (3) La₅₅Ni₁₀Al₃₅ rod 3 mm diameter ($T_g = 474$ K); and (4) Zr₆₂Cu₂₄Fe₅Al₉ rods, 1.5, 2.0 or 2.5 mm diameter ($T_g = 658$ K). Master alloys were prepared by arc-melting 3–4N pure elements. Glassy ribbons, 40 µm thick, were produced by melt-spinning, under argon atmosphere, on to a single copper wheel at 4,000 r.p.m. Bulk metallic glass (BMG) rods, with diameter 1.5–3.0 mm and length up to 80 mm, were cast, under Ti-gettered argon atmosphere, into water-cooled Cu-moulds, using induction melting and suction casting.

Thermal cycling. The La-based and Cu₄₆Zr₄₆Al₇Gd₁ samples were inserted into liquid nitrogen for 1 min, then heated by a hair-dryer set at room temperature, and held for 1 min; samples were treated with up to 25 such cycles. The Zr₆₂Cu₂₄Fe₅Al₉ samples were inserted into liquid nitrogen for 1 min, then into ethanol at 333–343 K for 1 min; samples were treated with up to 60 such cycles.

The paper reports only results from cycling between near room temperature and liquid-nitrogen temperature (77 K). Keeping the upper temperature fixed (near room temperature), it is expected that there would be an optimum lower temperature for maximum effect of thermal cycling (at given cooling and heating rates, and hold times at the upper and lower temperatures): for higher temperatures closer to room temperature, there would be less driving force for structural change; for lower temperatures further from room temperature, the rate of structural change near the lower temperature would be lower.

In addition to cycling down to 77 K, melt-spun ribbons of La₅₅Ni₂₀Al₂₅ glass were cycled down to dry-ice (CO₂) temperature (195 K) and to liquid-helium temperature (4.2 K) to explore the role of the lower temperature. The state of the samples was assessed using DSC (with heating rates from 20 to 40 K min⁻¹). For a given type of sample, the heat of relaxation ΔH_{rel} is lower for higher heating rate; it is therefore appropriate to normalize the ΔH_{rel} values relative to that for the as-cast glass measured at the same DSC heating rate. Measurements were made for samples: (1) held at 195 K for 100 min, for which the ratio of ΔH_{rel} to that of the as-cast glass is 1.23; (2) treated with ten cycles of 1 min at 195 K, for which the ratio is 1.24; (3) treated with ten cycles of 10 min at 195 K (ratio = 1.02); (4) treated with ten cycles of 1 min at 77 K (ratio = 1.55); and (5) treated with 13 cycles to 4.2 K (ratio is 1.37). Given the low number of samples tested, there is significant uncertainty in these relative values ($\pm 15\%$); nevertheless, it seems that, within the limited range of these tests, cycling down to 77 K produces the strongest effect.

A Ti-based BMG cycled up to 400 times between 77 K and 423 K (to simulate conditions on the surface of a spacecraft in low earth orbit) showed slight relaxation rather than rejuvenation³¹. This suggests that there is a significant annealing effect during the holds (25 min each) at the upper temperature, which itself is higher than in our work (and corresponds to $T/T_g = 0.68$, inside the range associated with β relaxation, see Fig. 1a in main text).

Stresses associated with thermal cycling. As noted in the paper, it is important to understand whether the effects of thermal cycling are due to the temperature changes themselves or to stresses arising from non-uniform temperature in the samples during cooling and heating. Whether the temperature is significantly non-uniform can be assessed using the dimensionless Biot number,

$$Bi = \frac{hL}{\kappa} \quad (1)$$

where h is the heat-transfer coefficient between the sample and the surrounding medium, L is a characteristic sample dimension (usually taken to be the ratio of sample volume to surface area), and κ is the thermal conductivity of the sample material. For a flat plate, L is the half-thickness; for a cylinder the half-radius. If $Bi < 0.1$, the temperature differences within the sample are negligible compared to the step in temperature between the sample and its surroundings (Newtonian cooling). Conversely, if $Bi > 1$, there are significant temperature gradients in the sample. We examine the conditions for samples immersed in liquid nitrogen.

Cooling in liquid nitrogen has been widely studied, particularly in connection with cryopreservation³². The ‘French straw’ sample containers used are cylinders of 2.8 mm diameter, and so are very similar in dimension to the cast BMG rods in the current work. On first immersion, there is film boiling, in which a film of vapour limits the heat transfer; in this regime³² $h = 148 \text{ W m}^{-2} \text{ s}^{-1}$. As the sample cools, there is a transition to the nucleate-boiling regime³² in which $h = 1,355 \text{ W m}^{-2} \text{ s}^{-1}$. The transition point has been measured for a variety of metallic samples

of geometry comparable to the samples in the present work³²; it occurs at temperatures ranging from 150 to 100 K. Thus a room-temperature sample immersed in liquid nitrogen spends around 75% of the temperature range down to 77 K with the relatively poor heat transfer associated with the film-boiling regime.

We used materials parameters determined for Vitreloy 1; Zr-based BMGs of this kind are among those for which thermal-cycling treatments might be put into practice. For Vitreloy 1, the thermal conductivity decreases with decreasing temperature³³ and starts to level off below room temperature. At 300 K, $\kappa = 4.59 \text{ W K}^{-1} \text{ m}^{-1}$; we use a rough estimate of $\kappa = 4 \text{ W K}^{-1} \text{ m}^{-1}$ for the temperature range in the present work. Extended Data Table 1 gives the estimates of Bi for the various sample geometries of interest.

We can see that for nearly all of these cases $Bi < 0.1$, and therefore temperature gradients (and thermal stresses) in the samples are negligible. Only for the largest-diameter cylinders in the final (nucleate boiling) stages of cooling is there any significant deviation from Newtonian cooling. These calculations support the experimental work (Extended Data Fig. 2 and associated discussion, see below) suggesting that sample size is not a significant factor in the thermal-cycling effect, and reinforce the conclusion that the effect is due to the temperature changes themselves and not due to any stresses arising from non-uniform temperature profiles in samples during cooling and heating.

It is of interest to consider the possible extent of thermal stresses associated with non-uniform sample temperature. Taking the value³⁴ determined for the Zr-based BMG Vitreloy 106, the linear coefficient of thermal expansion is $\alpha = 8.7 \times 10^{-6} \text{ K}^{-1}$. For a room-temperature sample immersed in liquid nitrogen, the maximum possible temperature difference during cooling is between 293 K (room temperature, possible at the sample centre) and 77 K (possible at the sample surface). For this temperature difference, the linear thermal strain is 0.19%, which would give internal stresses of the order of 10% of the macroscopic yield stress. In practice, as highlighted by the low values for the Biot number in Extended Data Table 1, any thermal stresses would be much lower than this value. Rejuvenation of metallic glasses by elastostatic loading²⁸ typically involves much higher stresses (up to 95% of the yield stress) for much longer times (tens of hours, rather than a few minutes). Again, it seems that stress effects in the current thermal cycling should be negligible.

Differential scanning calorimetry. This was performed (as in Fig. 2) using a Q2000 DSC (TA Instruments). Samples were heated at 20 K min⁻¹ from room temperature to 503 K (into the supercooled liquid state), held for 2 min, then cooled to room temperature at 20 K min⁻¹. A second cycle using the same procedure was used as the baseline for subtraction from the first cycle. The heat of relaxation ΔH_{rel} is calculated from the area between the two curves, from the onset of relaxation to the glass transition³⁵. Additional data were acquired using a Perkin Elmer DSC 8500 at a heating rate of 40 K min⁻¹.

Extended Data Figure 1a shows examples of DSC traces used to assess the reproducibility of measurements of the heat of relaxation ΔH_{rel} (using the baseline-subtraction method outlined above). Six samples of as-cast ribbons were tested, giving an average value of $\Delta H_{rel} = 740 \text{ J mol}^{-1}$, with a standard deviation of 30 J mol⁻¹ (that is, $\pm 4\%$).

As noted in the paper, it is important to know whether the effects of thermal cycling between near room temperature and 77 K are due just to the time spent at 77 K. Samples of La₅₅Ni₂₀Al₂₅ glassy ribbons were held at 77 K for 4 h (a time much longer than the total spent at 77 K in the course of ten room temperature–77 K cycles). Within experimental error (Extended Data Fig. 1b), such holds at 77 K appear to have no effect on ΔH_{rel} . Of course, a hold at 77 K represents a single cycle. As shown by the nano-indentation results in Fig. 3 in the main paper, even one cycle can have a measurable effect on properties. In the present case of ΔH_{rel} , however, any change as a result of one cycle is within the experimental uncertainty.

As noted in the main paper and discussed above in the Methods, it is necessary to examine whether the effects of thermal cycling are dependent on sample size. The non-uniform temperature in a larger sample could give rise to stresses and the cycling of those stresses could have effects. As-cast rods (3 mm diameter) of La₅₅Ni₁₀Al₃₅ BMG were subjected to room temperature–77 K cycles at the same time as thin (250–500 µm thick) discs pre-cut from the rods. After cycling, discs were cut from the centres of the bulk rods, so that samples of the same glass and same DSC-sample size could be compared after thermal cycling in two different geometries (thin disc and bulk). Extended Data Figure 2 shows that the changes in ΔH_{rel} are essentially the same for the samples thermally cycled in disc and in bulk form, suggesting that the stresses generated during cycling have a negligible effect. The two sets of data are combined in Fig. 2c.

Instrumented indentation. This was performed using an XP Nanoindenter (MTS Systems Corp.) at room temperature under load control up to a maximum load (F_{max}) of 40 mN. A diamond spherical indenter of tip radius 8.074 µm was used, its area function calibrated using fused silica and sapphire standards. Loading and unloading rates were 0.5 mN s⁻¹; the drift rate was maintained below

0.07 nm s^{-1} . Samples were loaded to F_{max} , then unloaded to $0.05F_{\text{max}}$ and maintained at that load for 60 s for measurement of thermal drift. The cumulative distribution curves (Fig. 3) include 39–48 data points. Initial yield events (pop-ins) were identified directly from the F – h curve and the associated spike in indenter tip velocity, or from the first deviation from the Hertzian elastic solution fitted to the F – h curve. Hardness was determined by the standard Oliver and Pharr method³⁶, on unloading at F_{max} .

Tests were conducted on melt-spun ribbons of $\text{La}_{55}\text{Ni}_{20}\text{Al}_{25}$ glass under the conditions noted above. Typical load-displacement curves (Extended Data Fig. 3) show pop-ins (that is, the transition from purely elastic to elastic-plastic deformation) with a moderate sharpness that is not affected by thermal cycling. The pop-ins are characterized by the values of initial yield load F_y and initial yield displacement Δh . Initial yield pressure P_y is calculated from:

$$P_y = \frac{F_y}{\pi a_y^2} \quad (2)$$

where a_y is the contact radius at initial yield. The reduced or indentation modulus E_r is determined by the standard method^{36,37}, on unloading from F_{max} . E_r is also determined by using the Hertzian elastic solution to fit the shape of the load-displacement curve up to the first pop-in. E_r is given by³⁸:

$$\frac{1}{E_r} = \frac{(1-\nu_i^2)}{E_i} + \frac{(1-\nu^2)}{E} \quad (3)$$

where E and ν are, respectively, the Young modulus and Poisson ratio of the glass, while the corresponding values³⁶ for the diamond indenter tip are taken to be $E_i = 1,141 \text{ GPa}$ and $\nu_i = 0.07$. Using equation (3), values of E were obtained.

Tests were conducted on ribbon samples in three states: (1) as-cast; (2) after one cycle from room temperature to a 10-min hold 77 K and back to room temperature; and (3) after a further ten cycles to 77 K each with 1 min hold. The effect of a longer hold at 77 K can thus be compared with cycling. All metallic-glass samples, as-cast or treated, are found to have some variation in properties over length scales of millimetres. For this reason it is important to compare data taken from a given region of the sample surface. In the present work, indentations were made in the same region (within 200 μm) of the given sample surface in the three successive states. Extended Data Table 2 collects data on the instrumented-indentation tests in the present work. Cumulative distributions, such as those shown in Fig. 3, are characterized by the median value and by the 1st and 9th deciles.

Extended Data Figure 4a shows the distribution of values of initial yield pressure P_y and initial yield displacement Δh for the three states of the $\text{La}_{55}\text{Ni}_{20}\text{Al}_{25}$ glass ribbons. On cycling, both P_y and Δh are reduced. There is a correlation, as expected, between higher values of P_y and higher values of Δh . The spread in values of Δh is particularly wide. After ten room temperature–77 K cycles, there is a marked reduction in the incidence of larger values: the initial yield events are smaller, reflecting less catastrophic shear banding.

The Young's modulus of the glass E determined by unloading from full load (Extended Data Fig. 4b) shows similar effects to those found for hardness H (Fig. 3b). The median value of E_{unload} in the as-cast sample decreases by 0.6% after 10 min hold at 77 K (compared to a 0.9% decrease in H), and by a further 7% after ten room temperature–77 K (1 min) cycles (compared to 4% for H). After the cycling treatments the relative width (1st to 9th decile) of the E distribution ($\pm 2.4\%$, compared to $\pm 3.1\%$ for H) is greater than in the as-cast sample ($\pm 1.4\%$, compared to $\pm 1.3\%$ for H).

The Young's modulus determined from the Hertzian fit of the F – h curve up to the first pop-in (Extended Data Fig. 4c) has a median value $\sim 5\%$ lower than that determined by unloading. The median value of E_{Hertz} in the as-cast sample decreases by 2.4% after 10-min hold at 77 K, and by a further 8.8% after ten room temperature–77 K (1 min) cycles. The relative width of the E_{Hertz} distribution in the as-cast sample ($\pm 1.6\%$) is similar to that for E_{unload} , but this width increases more (to $\pm 7.3\%$) after the cycling treatments. This suggests that the elastic properties up to the first pop-in are more sensitive to local variations in the sample.

Compression tests. These were made on Instron 5581 and Hounsfield 25 kN machines, loading at a strain rate of $5 \times 10^{-4} \text{ s}^{-1}$. The $\text{Zr}_{62}\text{Cu}_{24}\text{Fe}_5\text{Al}_9$ samples were cylindrical rods of 2:1 height:diameter, with diameter of 1.5, 2.0 or 2.5 mm. The $\text{Cu}_{46}\text{Zr}_{46}\text{Al}_7\text{Gd}_1$ samples were $1.5 \times 1.5 \times 3.0 \text{ mm}^3$ cuboids cut from the original cast rod of 3 mm diameter. All samples were polished to a mirror finish. The gradients of the elastic portions of the stress–strain curves in Fig. 4a are affected by the machine compliance and underestimate the Young's modulus. The curves are used only to extract values for the plastic strain as shown. The curves in Fig. 4c are almost fully corrected for machine compliance effects.

The curves in Fig. 4a and Fig. 4c show engineering stress and engineering strain. The strain values are calculated from the displacement of the testing machine and not from strain gauges on the samples. The gradients of the elastic portions of the curves are therefore affected by the machine compliance and significantly

underestimate the Young's modulus of the materials. The curves are used only to extract values for the plastic strain before catastrophic failure. Figure 4b collects data for rod samples of $\text{Zr}_{62}\text{Cu}_{24}\text{Fe}_5\text{Al}_9$ BMG of three diameters: 1.5 mm, 2.0 mm and 2.5 mm, and corresponding heights 3.0 mm, 4.0 mm and 5.0 mm. The stress–strain curves for 1.5 mm samples are shown in Fig. 4a; those for 2.0 mm and 2.5 mm samples in Extended Data Fig. 7.

As has been much discussed^{21,39}, larger samples are more susceptible to catastrophic failure through shear localization. Such size effects are extrinsic, and relate only to the mechanical tests themselves. In addition, there is an intrinsic effect: larger samples have a lower cooling rate during casting and so are more relaxed and tend to be more brittle in the as-cast state.

Resonant ultrasound spectroscopy. RUS measurements were conducted as in earlier work^{40,41}. Measurements were at room temperature on $1.5 \times 1.5 \times 3.0 \text{ mm}^3$ cuboids cut from the as-cast 3-mm-diameter rod of $\text{Cu}_{46}\text{Zr}_{46}\text{Al}_7\text{Gd}_1$ BMG. For each sample, a resonance spectrum was collected in the range 0.05–2.0 MHz with 65,000 data points. To detect changes in shear and bulk modulus, 10,000 data points were collected in each case in the ranges 0.900–0.910 MHz and 0.978–0.982 MHz, respectively. Within experimental error (less than $\pm 0.1\%$ for the shear modulus), room temperature–77 K thermal cycling has no effect on the elastic moduli of the glass (Extended Data Fig. 5). This suggests that the reductions in Young's modulus (Extended Data Fig. 4b, c) and correspondingly hardness (Fig. 3b) seen under quasi-static conditions are largely attributable to anelastic (that is, time-dependent elastic) strains.

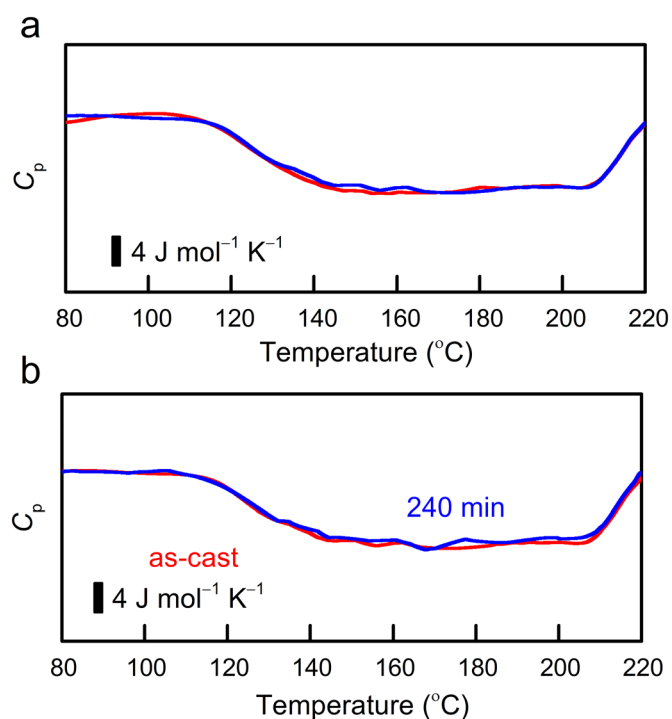
Dynamic mechanical analysis. This technique (not discussed in the paper), permitting measurement of the loss and storage moduli of a material, was applied (using methods reported previously⁴²) to ribbons of $\text{La}_{55}\text{Ni}_{20}\text{Al}_{25}$ glass (Extended Data Fig. 8) to explore the nature of the rejuvenation effect. The loss modulus shows a sharp maximum near the glass transition, associated with α relaxation. After ten room temperature–77 K cycles, the maximum is lowered by 2.3 K. This earlier onset of softening on heating is consistent with a rejuvenated state. The broad maximum at lower temperature (centred around 115 °C, Extended Data Fig. 8) shows β relaxation. The β relaxation in metallic glasses has been linked to mechanical properties^{42,43}. In view of the improved plasticity seen after thermal cycling (Fig. 4 and Extended Data Fig. 7), it is surprising that the effect of thermal cycling on the β relaxation is so slight (inset in Extended Data Fig. 8), especially compared to the clear effect on the α relaxation.

X-ray diffraction. There are reports that under cyclic mechanical loading at room temperature, in the elastic regime, a Zr-based BMG can undergo some crystallization⁴⁴. Deep cryogenic treatment of steels certainly induces phase changes^{22,25}. In the present work, X-ray diffraction (Bragg–Brentano geometry, Bruker D8 instrument, Cu K α radiation) was used to check that the samples remained fully glassy throughout, and especially after thermal cycling. Representative diffractograms for ribbon and bulk samples are shown in Extended Data Fig. 6. We found no evidence for any crystallization induced by the thermal-cycling treatments. The property changes suggest that the cycling does induce changes in the glassy structure but, typical of such structural relaxation, these changes are too subtle to be detectable with the simple X-ray methods used in the present work.

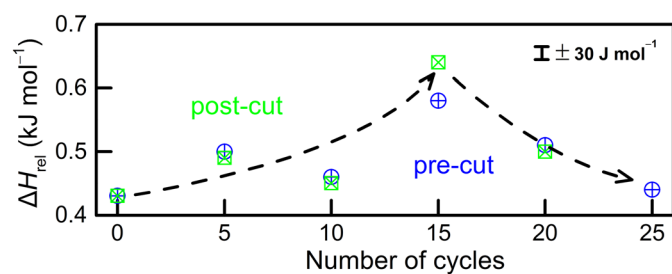
Other methods. Microhardness was measured on an Akashi MVK–HVL hardness testing machine, with 20 measurements for each data point. Scanning electron microscopy (secondary-electron imaging) was performed using a JEOL 5200 instrument.

- Wang, W. H. Roles of minor additions in formation and properties of bulk metallic glasses. *Prog. Mater. Sci.* **52**, 540–596 (2007).
- Wang, X., Shao, Y., Gong, P. & Yao, K. F. The effect of simulated thermal cycling on thermal and mechanical stability of a Ti-based bulk metallic glass. *J. Alloy. Comp.* **575**, 449–454 (2013).
- Victoria Santos, M., Sansinena, M., Chirife, J. & Zaritzky, N. Determination of heat transfer coefficients in plastic French straws plunged in liquid nitrogen. *Cryobiology* **69**, 488–495 (2014).
- Yamasaki, M., Kagao, S., Kawamura, Y. & Yoshimura, K. Thermal diffusivity and conductivity of supercooled liquid in $\text{Zr}_{41}\text{Ti}_{14}\text{Cu}_{12}\text{Ni}_{10}\text{Be}_{23}$ metallic glass. *Appl. Phys. Lett.* **84**, 4653–4655 (2004).
- Conner, R. D., Choi-Yim, H. & Johnson, W. L. Mechanical properties of $\text{Zr}_{57}\text{Nb}_{5}\text{Al}_{10}\text{Cu}_{15.4}\text{Ni}_{12.6}$ metallic glass matrix particulate composites. *J. Mater. Res.* **14**, 3292–3297 (1999).
- Mear, F. O., Lenk, B., Zhang, Y. & Greer, A. L. Structural relaxation in a heavily cold-worked metallic glass. *Scr. Mater.* **59**, 1243–1246 (2008).
- Oliver, W. C. & Pharr, G. M. An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. *J. Mater. Res.* **7**, 1564–1583 (1992).
- Field, J. S. & Swain, M. V. A simple predictive model for spherical indentation. *J. Mater. Res.* **8**, 297–306 (1993).
- Johnson, K. L. *Contact Mechanics* (Cambridge Univ. Press, 1985).
- Cheng, Y. Q., Han, Z., Li, Y. & Ma, E. Cold versus hot shear banding in bulk metallic glass. *Phys. Rev. B* **80**, 134115 (2009).

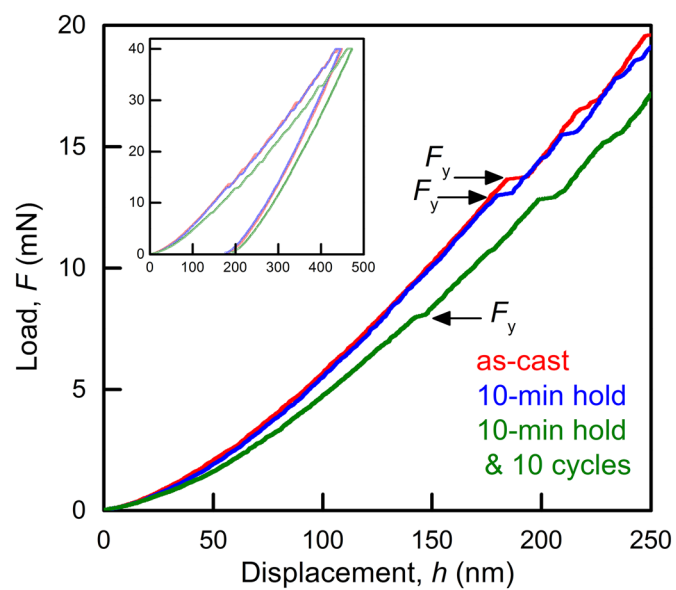
40. McKnight, R. E. A. *et al.* Grain size dependence of elastic anomalies accompanying the alpha-beta phase transition in polycrystalline quartz. *J. Phys. Condens. Matter* **20**, 075229 (2008).
41. Concustell, A., Godard-Desmarest, S., Carpenter, M. A., Nishiyama, N. & Greer, A. L. Induced elastic anisotropy in a bulk metallic glass. *Scr. Mater.* **64**, 1091–1094 (2011).
42. Yu, H. B. *et al.* Tensile plasticity in metallic glasses with pronounced β relaxations. *Phys. Rev. Lett.* **108**, 015504 (2012).
43. Yu, H. B., Samwer, K., Wang, W. H. & Bai, H. Y. Chemical influence on β -relaxations and the formation of molecule-like metallic glasses. *Nature Commun.* **4**, 2204 (2013).
44. Churyumov, A. & Yu. . *et al.* Phase transformations in Zr-based bulk metallic glass cyclically loaded before plastic yielding. *Mater. Sci. Eng. A* **550**, 358–362 (2012).



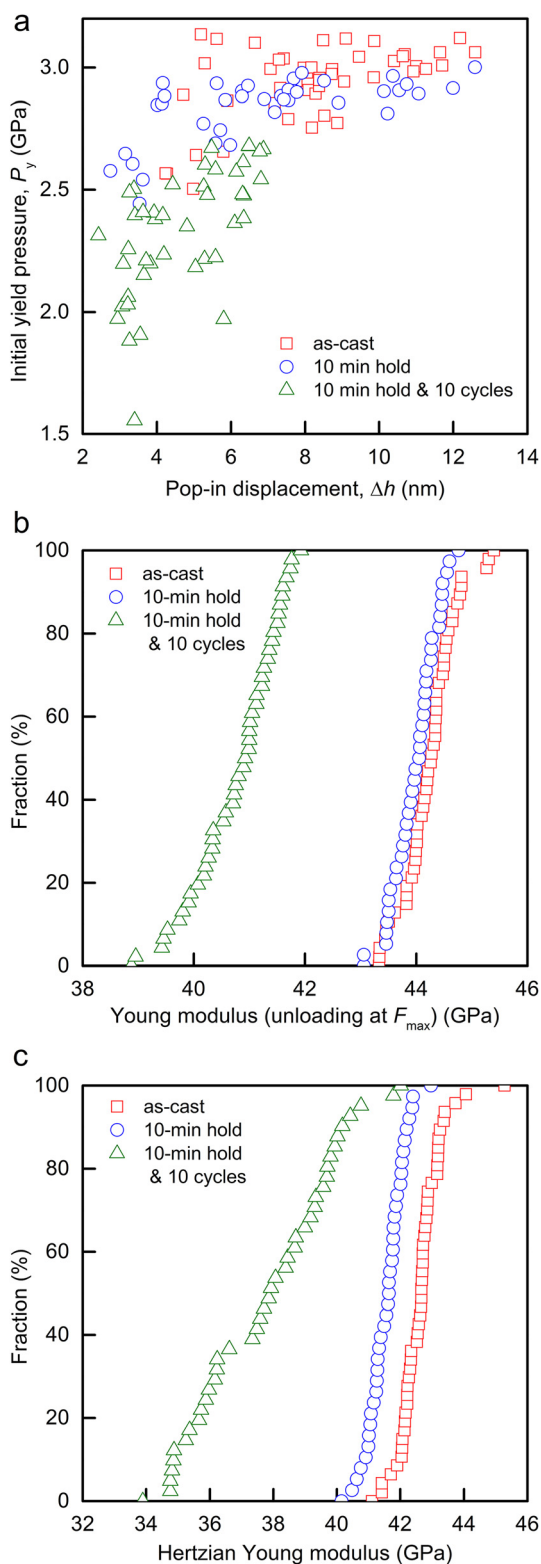
Extended Data Figure 1 | DSC traces for ribbons of $\text{La}_{55}\text{Ni}_{20}\text{Al}_{25}$ glass (heated at 20 K min^{-1}). **a**, These example traces for as-cast samples show the exotherm, just below the glass-transition temperature, from which the heat of relaxation ΔH_{rel} is determined. **b**, The effect of holding at 77 K: the traces show that the heat of relaxation ΔH_{rel} , given by the exotherm just below the glass-transition temperature, is, within the experimental error of 4%, the same for an as-cast ribbon, and for a sample held for 4 h at 77 K.



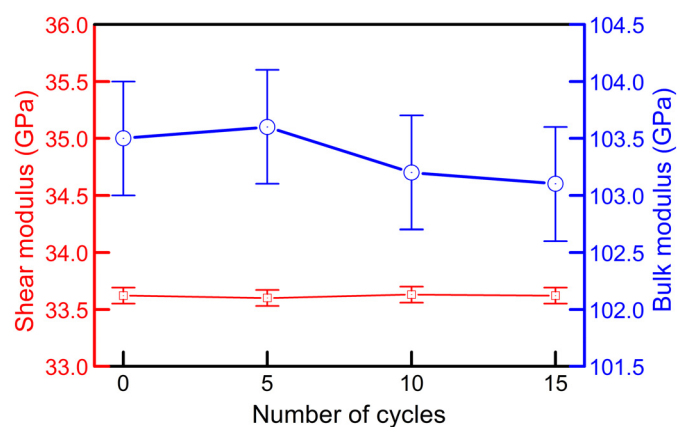
Extended Data Figure 2 | Differential scanning calorimetry (DSC) of $\text{La}_{55}\text{Ni}_{10}\text{Al}_{35}$ bulk metallic glass. The heat of relaxation ΔH_{rel} is compared for discs (250–500 μm thick) pre-cut from the bulk rod and then treated with room temperature–77 K cycles, and for discs post-cut from treated rod samples. There is no difference between these cases within the error of $\pm 30 \text{ J mol}^{-1}$.



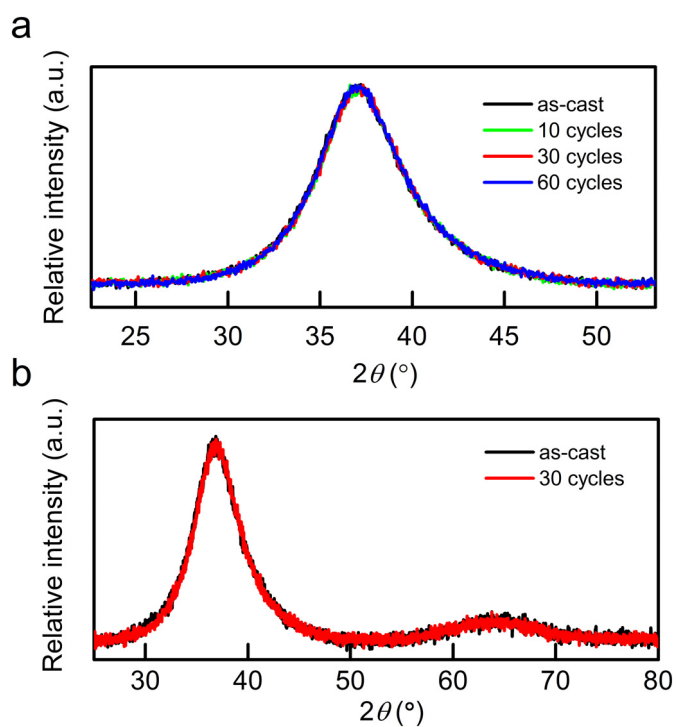
Extended Data Figure 3 | Load-displacement curves for ribbons of $\text{La}_{55}\text{Ni}_{20}\text{Al}_{25}$ metallic glass tested up to a maximum load F_{max} of 40 mN. The initial yield load F_y is indicated on the curves for the glass in three states: as-cast, after a 10-min hold at 77 K, and after a further ten room temperature–77 K cycles each with 1 min hold.



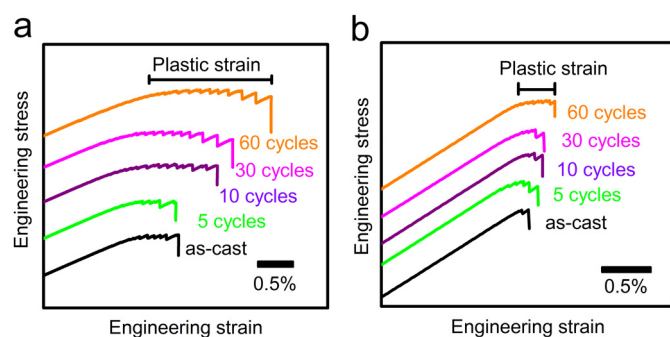
Extended Data Figure 4 | Instrumented indentation of $\text{La}_{55}\text{Ni}_{20}\text{Al}_{25}$ metallic glass ribbon. **a**, Initial yielding, characterized by initial yield pressure P_y and initial yield displacement Δh . Indentations are made for three states of the sample: as-cast, after a 10-min hold at 77 K, and after a further ten room temperature–77 K cycles each with 1-min hold. **b**, **c**, Distributions of the Young's modulus, using data from the same indentations of the sample in three states as in Fig. 3 and in **a**. Values of the Young's modulus in the glass E are determined (**b**) by the standard Oliver and Pharr³⁶ method on unloading from F_{max} and (**c**) from Hertzian fitting to the F – h curve up to the first pop-in.



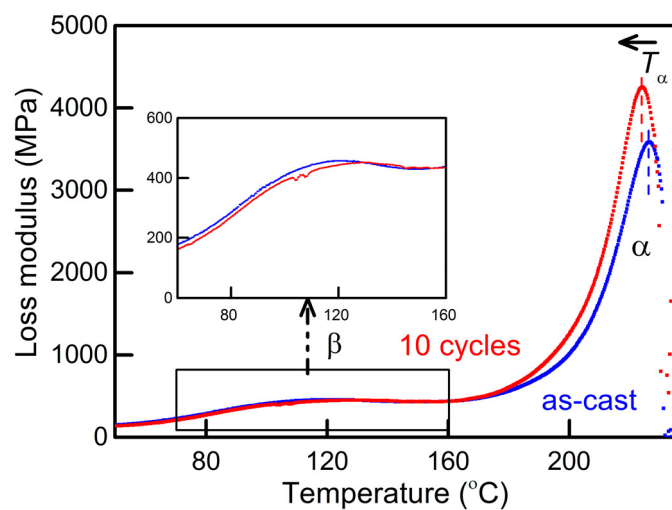
Extended Data Figure 5 | Elastic moduli derived from resonant ultrasound spectroscopy for $\text{Cu}_{46}\text{Zr}_{46}\text{Al}_7\text{Gd}_1$ BMG treated with room temperature–77 K thermal cycles. Shear modulus (left axis); bulk modulus (right axis). Error bars, root mean square errors in the fitted frequencies.



Extended Data Figure 6 | X-ray diffraction traces for Zr-based metallic glasses subjected to 338–77 K thermal cycles. a, For melt-spun ribbons of $\text{Zr}_{61.1}\text{Cu}_{26.3}\text{Fe}_{2.1}\text{Al}_{10.5}$; **b,** for rods of $\text{Zr}_{62.2}\text{Cu}_{23.9}\text{Fe}_{4.8}\text{Al}_{9.1}$. No clear changes are induced by the cycling treatments.



Extended Data Figure 7 | Compressive stress–strain curves for rod samples of $\text{Zr}_{62}\text{Cu}_{24}\text{Fe}_5\text{Al}_9$ BMG. **a**, For rods of 2 mm diameter; **b**, for rods of 2.5 mm diameter. In each case, increasing numbers of 338 K to 77 K thermal cycles cause the plastic strain to increase.



Extended Data Figure 8 | Dynamic mechanical analysis of ribbons of $\text{La}_{55}\text{Ni}_{20}\text{Al}_{25}$ metallic glass. The heating rate is 3 K min^{-1} . The general form of the curve matches that shown in Fig. 1a, where an example was chosen of a metallic glass showing β relaxation at a particularly low value of T/T_g . For $\text{La}_{55}\text{Ni}_{20}\text{Al}_{25}$ the β relaxation is centred at $T/T_g \approx 0.8$.

Extended Data Table 1 | Values of the Biot number Bi for immersion in liquid nitrogen of the various sample geometries in the present work

Sample geometry	L (μm)	Bi , film boiling	Bi , nucleate boiling
melt-spun ribbon, 40 μm thick	20	7.4×10^{-4}	6.8×10^{-3}
disc, 250–500 μm thick	125–250	$(4.6\text{--}9.2) \times 10^{-3}$	$(4.2\text{--}8.5) \times 10^{-2}$
cylinder, 1.5 mm diam.	375	1.4×10^{-2}	0.13
cylinder, 2.0 mm diam.	500	1.8×10^{-2}	0.17
cylinder, 2.5 mm diam.	625	2.3×10^{-2}	0.21
cylinder, 3.0 mm diam.	750	2.8×10^{-2}	0.26

The film-boiling regime applies until the sample surface temperature decreases to between 150 K and 100 K, and the nucleate-boiling regime applies thereafter.

Extended Data Table 2 | Summary of instrumented-indentation tests

State Property	as-cast			10 min LN			10 min & 10 cycles LN		
	10%	median	90%	10%	median	90%	10%	median	90%
No. of indents	48			39			47		
F_y (mN) \pm 0.1	10.6	14.3	16.0	10.6	13.5	14.3	5.8	9.3	11.5
Δh (nm) \pm 0.2	5.1	8.2	11.4	3.6	6.8	10.7	3.2	4.5	6.5
P_y (GPa) \pm 0.015	2.69	2.98	3.12	2.61	2.89	2.96	1.97	2.38	2.66
h_{max} (nm) \pm 1	442	445	450	444	449	453	458	469	481
H (GPa) \pm 0.003	2.607	2.647	2.678	2.576	2.622	2.672	2.450	2.518	2.607
E_{unload} (GPa) \pm 0.04	43.52	44.29	44.79	43.47	44.04	44.47	39.65	40.93	41.59
E_{Hertz} (GPa) \pm 0.04	41.95	42.65	43.33	40.83	41.64	42.22	34.84	37.90	40.35

Cumulative distributions, such as those shown in Fig. 3, are characterized by the median value and by the first and ninth deciles.

Graphene kirigami

Melina K. Bles¹, Arthur W. Barnard², Peter A. Rose¹, Samantha P. Roberts¹, Kathryn L. McGill¹, Pinshane Y. Huang², Alexander R. Ruyack³, Joshua W. Kevek¹, Bryce Kobrin¹, David A. Muller^{2,4} & Paul L. McEuen^{1,4}

For centuries, practitioners of origami ('ori', fold; 'kami', paper) and kirigami ('kiri', cut) have fashioned sheets of paper into beautiful and complex three-dimensional structures. Both techniques are scalable, and scientists and engineers are adapting them to different two-dimensional starting materials to create structures from the macro- to the microscale^{1,2}. Here we show that graphene^{3–6} is well suited for kirigami, allowing us to build robust microscale structures with tunable mechanical properties. The material parameter crucial for kirigami is the Föppl-von Kármán number^{7,8} γ : an indication of the ratio between in-plane stiffness and out-of-plane bending stiffness, with high numbers corresponding to membranes that more easily bend and crumple than they stretch and shear. To determine γ , we measure the bending stiffness of graphene monolayers that are 10–100 micrometres in size and obtain a value that is thousands of times higher than the predicted atomic-scale bending stiffness. Interferometric imaging attributes this finding to ripples in the membrane^{9–13} that stiffen the graphene sheets considerably, to the extent that

γ is comparable to that of a standard piece of paper. We may therefore apply ideas from kirigami to graphene sheets to build mechanical metamaterials such as stretchable electrodes, springs, and hinges. These results establish graphene kirigami as a simple yet powerful and customizable approach for fashioning one-atom-thick graphene sheets into resilient and movable parts with microscale dimensions.

Devices such as those shown in Fig. 1a, c, d are made from polycrystalline monolayer graphene that is grown on copper by chemical vapour deposition¹⁴, and then transferred to fused silica wafers that are covered with an aluminium release layer. We use optical lithography to pattern both the graphene and the 50-nm-thick gold pads that are deposited on top of the graphene to act as handles. Finally, we release the graphene from the surface by etching away the aluminium in mild acid. The devices remain in aqueous solution with added salts or surfactants as desired. An inverted white-light microscope with a video camera is used to image the sheets, and micromanipulators are used to probe them.

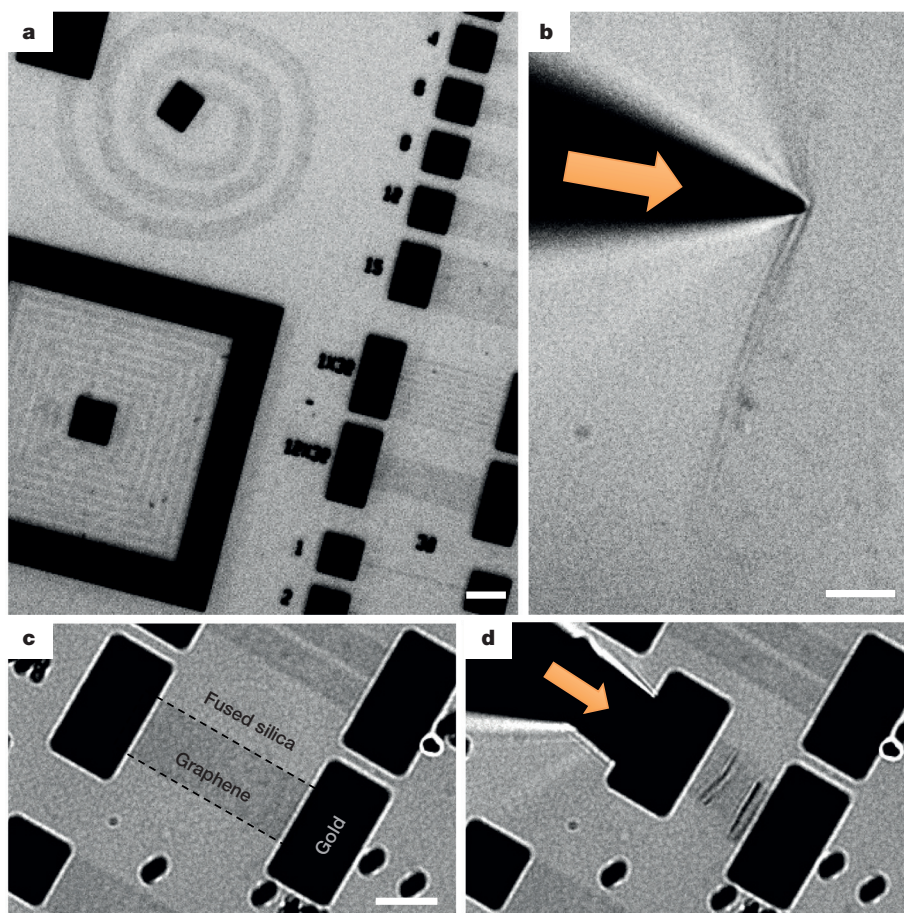


Figure 1 | Patterning and manipulating graphene. **a**, Transmission white-light image showing completed devices: a spiral spring, a kirigami pyramid, and a variety of cantilevers. **b**, Manipulating a large sheet of graphene with a micromanipulator. The sheet folds and crumples like soft paper, and returns to its original shape. **c**, **d**, Manipulating devices with gold pads. The devices can be lifted entirely off the surface; see Supplementary Video 1. Scale bars are 10 μm . All images and videos have undergone linear contrast adjustments.

¹Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA. ²School of Applied and Engineering Physics, Cornell University, Ithaca, New York 14853, USA. ³School of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14853, USA. ⁴Kavli Institute at Cornell for Nanoscale Science, Cornell University, Ithaca, New York 14853, USA.

We move the graphene along the surface or peel it up entirely by pushing a sharp probe tip into the gold pads or against the graphene itself (Fig. 1b–d and Supplementary Video 1). The graphene's elastic behaviour is reminiscent of that of thin paper: it folds and crumples out of plane, but does not notably stretch in plane (Fig. 1b). The process is almost entirely reversible in the presence of surfactants, even after considerable crumpling of the graphene.

The mechanical properties relevant for kirigami are captured by the Föppl–von Kármán number^{7,8} for a square sheet of side length L and thickness t : $\gamma = Y_{2D}L^2/\kappa \approx (L/t)^2$, that is, the ratio between the two-dimensional Young's modulus Y_{2D} and the out-of-plane bending stiffness κ , multiplied by the length squared. To determine γ , we measure κ by using the photon pressure from an infrared laser to apply a known force to a pad attached to a graphene cantilever and measuring the resulting displacement (Fig. 2a). We also measure thermal fluctuations of cantilevers to determine their spring constants (Fig. 2b and Extended Data Fig. 4), which, according to the equipartition theorem, are $k = k_B T / \langle x_{th}^2 \rangle$, where T is temperature, k_B is Boltzmann's constant, and $\langle x_{th}^2 \rangle$ is the time-averaged square of the cantilever thermal fluctuation amplitude. Although the presence of water (the aqueous solution in which the device is immersed) slows down the fluctuations, it does not change the spring constant^{15,16}. Cantilevers with lengths of 8–80 μm and widths of 2–15 μm have spring constants of 10^{-5} – 10^{-8} N m^{-1} . These are astonishingly soft springs, as many as eight orders of magnitude softer than a typical atomic force microscope cantilever. The bending stiffness κ is inferred from the measured spring constant using $k = 3\kappa W/L^3$, where W and L are the width

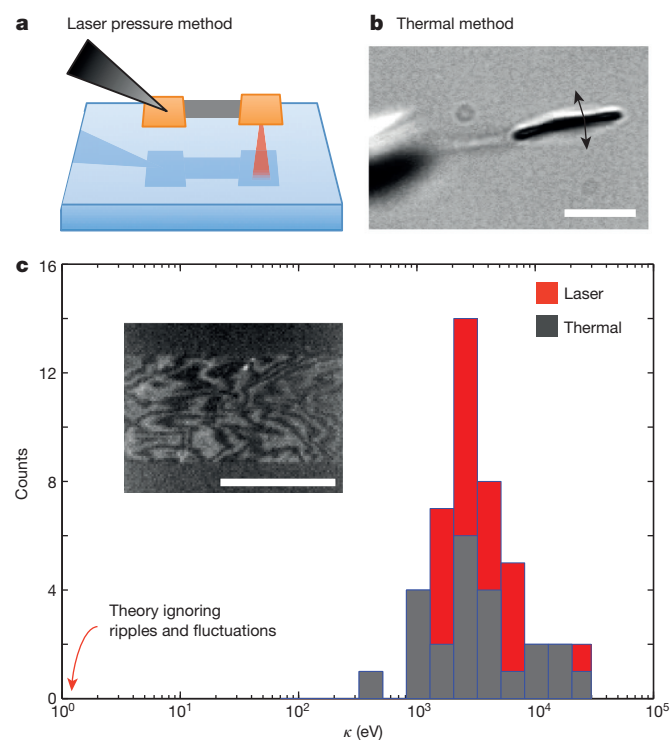


Figure 2 | Measuring the bending stiffness of monolayer graphene.

a, Applying controlled forces to a gold pad using an infrared laser. The grey triangle represents the probe tip that holds the device up off the surface; the red triangle represents the focused laser beam. The cantilever displacement gives the spring constant. **b**, Tracking the motion of a rotated device under thermal fluctuations provides an independent measurement of the spring constant (Extended Data Fig. 3). **c**, Stacked histogram of bending stiffness, and interference micrographs of devices whose aluminium release layer has been etched away, showing the structure of static ripples (inset). The spring constant relates to the bending stiffness as $k = 3\kappa W/L^3$. The red arrow points to the microscopic bending stiffness, $\kappa_0 = 1.2 \text{ eV}$. Scale bars are 10 μm . Interference images were averaged over 180 frames at 90 frames per second.

and length of the cantilevers, respectively. The values obtained from these thermal measurements and the laser measurements are shown in Fig. 2c, and are seen to be orders of magnitude higher than $\kappa_0 = 1.2 \text{ eV}$, which is the value that is predicted from the microscopic bending stiffness of graphene (known from simulations¹⁷ and measurements of the phonon modes in graphite¹⁸).

Both thermal fluctuations and static ripples are predicted to notably stiffen ultrathin crystalline membranes^{9–13,19,20} by effectively thickening the membrane, similar to how a crumpled sheet of paper is more rigid than a flat one. For static ripples, the effective bending stiffness is predicted to be⁹ $\kappa_{\text{eff}}/\kappa_0 \approx \sqrt{Y_{2D}\langle z_{\text{eff}}^2 \rangle}/\kappa_0$, where $\langle z_{\text{eff}}^2 \rangle$ is the space-averaged square of the effective amplitude of the static ripples and $Y_{2D} = 340 \text{ N m}^{-1}$ is the two-dimensional Young's modulus³. For an initially flat membrane with thermal fluctuations, the stiffness is predicted to be $\kappa_{\text{eff}} \approx \kappa_0(W/l_c)^\eta$, where $l_c = \sqrt{32\pi^3\kappa_0^2/(3Y_{2D}k_B T)}$ is the Ginzburg length¹⁹, and η is a scaling exponent.

We look for static ripples in graphene cantilevers using interference microscopy²¹ (inset of Fig. 2c). The black bands in such images are regions of constant elevation, with the spacing between black and white bands corresponding to changes in z of $\lambda/4 \approx 100 \text{ nm}$ (where λ is the wavelength, corrected for the refractive index of water). With a typical $\langle z_{\text{eff}}^2 \rangle$ value from these measurements of about $(100 \text{ nm})^2$, we obtain an effective bending stiffness of $\kappa_{\text{eff}}/\kappa_0 \approx 4,000$.

Static ripples are present only after releasing graphene from the surface (Extended Data Fig. 2), and likely to be sample specific and influenced by growth, fabrication details, and so on. Developing growth and fabrication protocols that can change the amplitude of the static ripples or eliminate them altogether is of great interest. Other groups have observed ripples in suspended (strained) graphene membranes^{5,22}, although they occur at a much smaller scale and their origin remains a subject of debate. Moreover, the thermal theory outlined above predicts a bending stiffness at room temperature due to thermal fluctuations of $\kappa_{\text{eff}}/\kappa_0 \approx 1,000$ for an initially flat membrane. These contradictory findings call for future experiments to firmly establish the relative contribution to bending stiffness of thermal fluctuations and static ripples²³. But irrespective of cause, the high bending stiffness notably changes the effective γ value, $\gamma_{\text{eff}} = Y_{\text{eff}}L^2/\kappa_{\text{eff}}$. With the predicted renormalization^{9,11} of Y_{eff} , we find that γ_{eff} is of the order of 10^5 – 10^7 for a sheet of graphene $10 \mu\text{m} \times 10 \mu\text{m}$ in size, close to that of a standard sheet of paper.

The mechanical similarity between graphene and paper makes it easy to translate ideas and intuition directly from paper models to graphene devices. For example, the highly stretchable graphene transistors in Fig. 3b, c and Supplementary Video 2 are based on a simple kirigami pattern of alternating, offset cuts and are created using photolithography. Here, the elasticity of the kirigami spring is determined by the pattern of cuts and the bending stiffness (rather than the Young's modulus) of the graphene. As the reconstruction of the three-dimensional shape of a stretched and lifted device in Fig. 3e shows, the graphene strips pop up and bend out of plane as the spring is stretched.

We measure the electrical response of these stretchable transistors by gating them with an approximately 10 mM KCl solution²⁴ (see Methods for details). Figure 3d plots the liquid-gate dependence of the conductance at a source–drain bias of 100 mV for a device in its initial unstretched state (blue) and when stretched by 240% (orange). The normalized change in conductance with gate voltage per graphene square is 0.7 mS V^{-1} and the resistance per graphene square at the Dirac point is 12 k Ω , comparable to what has been reported for electrolyte-gated graphene transistors²⁴. Because the graphene lattice itself is not much strained when the kirigami spring is extended, we do not expect or observe a notable change in the conductance curves between the unstretched and stretched states, which is highly desirable for stretchable electronics²⁵. Furthermore, stretching and unstretching a similar device more than 1,000 times did not substantially change its electrical properties.

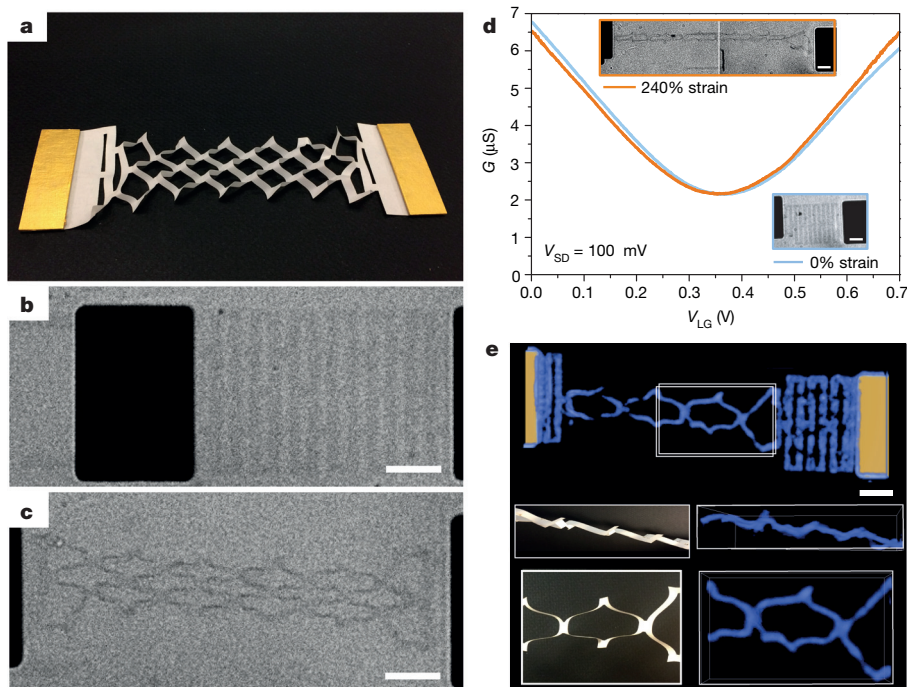


Figure 3 | Stretchable graphene transistors.

a, b, Paper and graphene in-plane kirigami springs, respectively. **c,** Graphene spring stretched by about 70%. **d,** Electrical properties in approximately 10 mM KCl. Conductance G is plotted against liquid-gate voltage V_{LG} at source-drain bias $V_{SD} = 100$ mV before stretching (blue) and when stretched by 240% (orange). The top (orange-boxed) inset is split because the stretched device was larger than the visible area. **e,** Three-dimensional reconstruction from a z-scan focal series of a graphene spring. The right side remains stuck to the surface and the left side is lifted. Insets show views of sections of the graphene (right) and paper models (left). Top images show side views; bottom images show top views. The thin grey lines are the bounding box from the three-dimensional reconstruction. The aspect ratio of the side-view paper model was compressed 1.8×. Scale bars are 10 μm. See Supplementary Video 2.

Figure 4a and Supplementary Video 3 show graphene cut so that it forms an out-of-plane pyramidal spring, along with a paper model. The spring's force-distance curve in Fig. 4b, measured with the photon pressure from an infrared laser focused on the central pad, gives a spring constant of $k = 2 \times 10^{-6} \text{ N m}^{-1}$. This value compares well with the spring constant estimate of (5×10^{-7}) – $(5 \times 10^{-6}) \text{ N m}^{-1}$, obtained from our κ measurements (Fig. 2) and the geometry of the device.

Remote actuation of the kirigami devices is possible using magnetic fields or linked graphene elements. Figure 4c illustrates magnetic actuation, with magnetic forces and torques acting on an attached iron pad allowing for parallel (many hinges being controlled simultaneously) and complex manipulations to be made. The opening and closing of the graphene hinge (which is 1 μm long and 10 μm wide) in Fig. 4d uses a longer graphene strip (out of focus in the image) that extends in

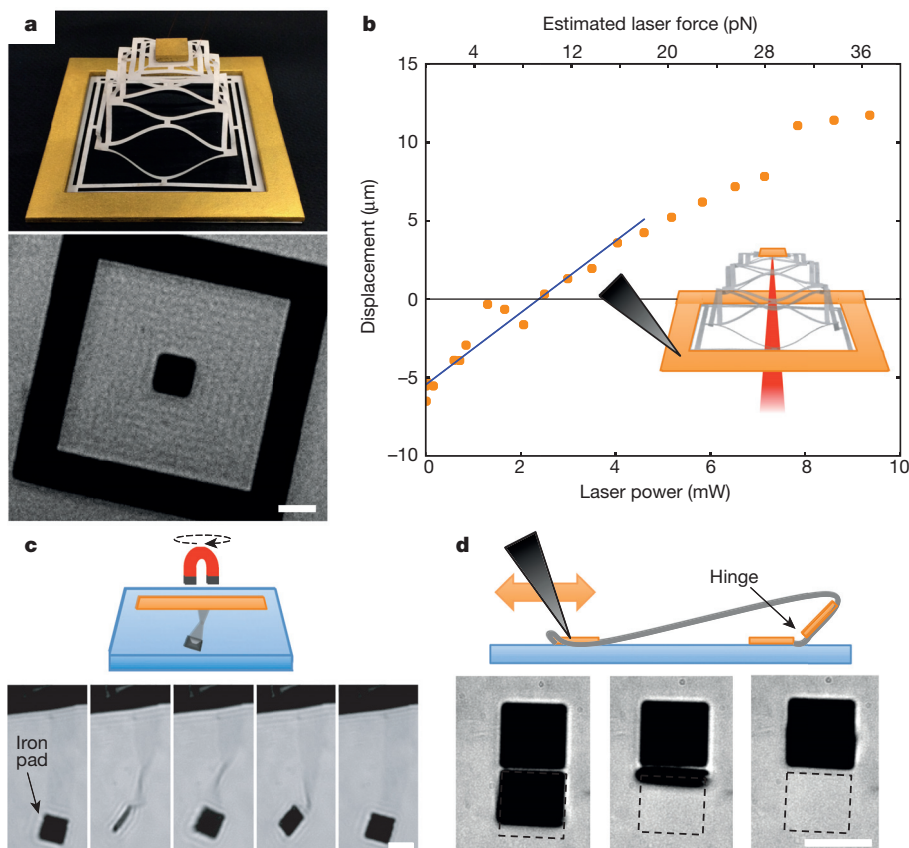


Figure 4 | Remote actuation. **a,** Paper model (top) and as-fabricated graphene kirigami pyramid (bottom). We actuate this out-of-plane spring using an infrared laser. **b,** Schematic and force-distance curve for a pyramid such as the one shown in **a**. A linear fit at low forces yields $k = 2 \times 10^{-6} \text{ N m}^{-1}$. **c,** A rotating static magnetic field twists and untwists a long strip of graphene. The gold pad is replaced by iron. **d,** A monolayer graphene hinge actuated by a graphene arm. Stills show the hinge closing. Supplementary Video 3 was taken after opening and closing this hinge 1,000 times; it survived 10,000 cycles. Scale bars are 10 μm.

a loop over the hinge to a gold pad, so that moving this probe along the surface opens and closes the hinge (Supplementary Video 3). Although the hinge is only one atom thick, it survived more than 10,000 open-and-close cycles (at which point the gold pads started to warp). This remarkable resilience and the scope for scaling down to tens of nanometres make monolayer graphene hinges ideally suited for use in microscale moving parts.

We envisage that graphene kirigami will have many useful applications. For example, springs like those in Figs 3 and 4 are easily designed, with spring constants that range from 1 N m^{-1} to 10^{-8} N m^{-1} (which covers the full range from atomic force microscopes to optical traps), for use as force measurement devices with a simple visual readout and femtonewton force resolution. The addition of elements such as bimorphs²⁶ or chemical tags to graphene kirigami devices²⁷ will create environment-responsive metamaterials. Kirigami techniques can also easily be applied to other two-dimensional materials that have different optical, electronic, and mechanical properties, which creates opportunities for further development of self-actuated two-dimensional functional devices that respond to light or magnetic fields, changes in temperature, or chemical signals. Such atomically thin membrane devices may be used for sensing, manipulation, complex origami, and nanoscale robotics.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 March; accepted 14 May 2015.

Published online 29 July 2015.

- Wang-Iverson, P., Lang, R. J. & Yim, M. (eds) *Origami 5: Fifth International Meeting of Origami Science, Mathematics, and Education* (CRC Press, 2011).
- Hawkes, E. *et al.* Programmable matter by folding. *Proc. Natl Acad. Sci. USA* **107**, 12441–12445 (2010).
- Lee, C., Wei, X., Kysar, J. W. & Hone, J. Measurement of the elastic properties and intrinsic strength of monolayer graphene. *Science* **321**, 385–388 (2008).
- Booth, T. J. *et al.* Macroscopic graphene membranes and their extraordinary stiffness. *Nano Lett.* **8**, 2442–2446 (2008).
- Meyer, J. C. *et al.* The structure of suspended graphene sheets. *Nature* **446**, 60–63 (2007).
- Bunch, J. S. *et al.* Electromechanical resonators from graphene sheets. *Science* **315**, 490–493 (2007).
- Föppl, A. *Vorlesungen über technische Mechanik* (B. G. Teubner, 1905).
- von Kármán, T. *Festigkeitsproblem im Maschinenbau*. Vol. 4 (Encyklopädie der Mathematischen Wissenschaften, 1910).
- Košmrlj, A. & Nelson, D. R. Mechanical properties of warped membranes. *Phys. Rev. E* **88**, 012136 (2013).
- Nelson, D. R. & Peliti, L. Fluctuations in membranes with crystalline and hexatic order. *J. Phys.* **48**, 1085–1092 (1987).
- Aronovitz, J. A. & Lubensky, T. C. Fluctuations of solid membranes. *Phys. Rev. Lett.* **60**, 2634–2637 (1988).
- Le Doussal, P. & Radzihovsky, L. Self-consistent theory of polymerized membranes. *Phys. Rev. Lett.* **69**, 1209–1212 (1992).
- Los, J. H., Katsnelson, M. I., Yazyev, O. V., Zakharchenko, K. V. & Fasolino, A. Scaling properties of flexible membranes from atomistic simulations: application to graphene. *Phys. Rev. B* **80**, 121405 (2009).
- Li, X. *et al.* Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science* **324**, 1312–1314 (2009).
- Velasco, S. On the Brownian motion of a harmonically bound particle and the theory of a Wiener process. *Eur. J. Phys.* **6**, 259–265 (1985).
- te Velthuis, A. J. W., Kerssemakers, J. W. J., Lipfert, J. & Dekker, N. H. Quantitative guidelines for force calibration through spectral analysis of magnetic tweezers data. *Biophys. J.* **99**, 1292–1302 (2010).
- Fasolino, A., Los, J. H. & Katsnelson, M. I. Intrinsic ripples in graphene. *Nature Mater.* **6**, 858–861 (2007).
- Nicklow, R., Wakabayashi, N. & Smith, H. G. Lattice dynamics of pyrolytic graphite. *Phys. Rev. B* **5**, 4951–4962 (1972).
- Roldán, R., Fasolino, A., Zakharchenko, K. V. & Katsnelson, M. I. Suppression of anharmonicities in crystalline membranes by external strain. *Phys. Rev. B* **83**, 174104 (2011).
- Braghin, F. L. & Hasselmann, N. Thermal fluctuations of free-standing graphene. *Phys. Rev. B* **82**, 035407 (2010).
- Georgiou, T. *et al.* Graphene bubbles with controllable curvature. *Appl. Phys. Lett.* **99**, 093103 (2011).
- Wang, W. L. *et al.* Direct imaging of atomic-scale ripples in few-layer graphene. *Nano Lett.* **12**, 2278–2282 (2012).
- Košmrlj, A. & Nelson, D. R. Thermal excitations of warped membranes. *Phys. Rev. E* **89**, 022126 (2014).
- Chen, D., Tang, L. & Li, J. Graphene-based materials in electrochemistry. *Chem. Soc. Rev.* **39**, 3157–3180 (2010).
- Rogers, J. A., Someya, T. & Huang, Y. Materials and mechanics for stretchable electronics. *Science* **327**, 1603–1607 (2010).
- Zhu, S.-E. *et al.* Graphene-based bimorph microactuators. *Nano Lett.* **11**, 977–981 (2011).
- Yuk, J. M. *et al.* Graphene veils and sandwiches. *Nano Lett.* **11**, 3290–3294 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Nelson, M. Bowick, A. Kosmrlj, J. Alden, A. van der Zande, and R. Martin-Wells for discussions. We thank E. Minot for assistance with electrolyte gating. We thank R. Hovden for discussions on three-dimensional reconstruction theory and techniques, and R. Hovden, M. Hanwell, and U. Ayachit for developing and supporting the TomViz three-dimensional visualization software. We thank J. Wardini, P. Ong, A. Zaretski, and S. P. Wang for additional graphene samples, and F. Parish with Cornell's College of Architecture, Art, and Planning for assistance with the paper models. We also acknowledge the Origami Resource Center (<http://www.origami-resource-center.com/>) for kirigami design ideas. This work was supported by the Cornell Center for Materials Research (National Science Foundation, NSF, grant DMR-1120296), the Office of Naval Research (N00014-13-1-0749), and the Kavli Institute at Cornell for Nanoscale Science. Devices were fabricated at the Cornell Nanoscale Science and Technology Facility, a member of the National Nanotechnology Infrastructure Network, which is supported by the NSF (ECCS-0335765). K.L.M. and P.Y.H. acknowledge support from the NSF Graduate Research Fellowship Program (DGE-1144153 and DGE-0707428). Tomography visualization software development was supported by a SBIR grant (DE-SC0011385).

Author Contributions Device design and actuation techniques were developed by M.K.B., A.W.B., P.A.R., S.P.R., and K.L.M. under the supervision of P.L.M. Fabrication and characterization was performed by the above authors with additional support from A.R.R., J.W.K., and B.K. Bending stiffness measurements were designed by A.W.B. and P.L.M. and carried out by M.K.B., P.A.R., and K.L.M. with data analysis by S.P.R., A.W.B., M.K.B., K.L.M., and P.A.R. under the supervision of P.L.M. Electrical measurements were performed by K.L.M. and M.K.B. under the supervision of P.L.M. Three-dimensional reconstructions were performed by P.Y.H. under the supervision of D.A.M. The paper was written by M.K.B. and P.L.M., with assistance from P.A.R., A.W.B., K.L.M., and P.Y.H. and in consultation with all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.L.M. (plm23@cornell.edu).

METHODS

Graphene growth. We grow graphene on copper following a standard chemical vapour deposition process¹⁴. The copper foil is purchased from Alpha Aesar, stock number 13382. The copper is annealed for 36 min at 980 °C with a H₂ flow of 60 standard cubic centimetres per minute (s.c.c.m.). Graphene is grown at 980 °C for 20 min with a H₂ flow of 60 s.c.c.m. and a CH₄ flow of 36 s.c.c.m. The foil is then cooled in a matching environment as quickly as possible.

Graphene characterization. Typical Raman spectra, scanning electron microscope images, and bright-field transmission electron microscope (TEM) images all confirm that the growths yielded mostly single-layer graphene with small bilayer regions (Extended Data Fig. 1). Dark-field TEM on a variety of growths reveal that typical grain sizes are of the order of hundreds of nanometres to micrometres.

Fabrication of cantilevers and kirigami devices. Fabrication follows standard graphene processing methods, with the addition of an aluminium release layer. We evaporate 40 nm of aluminium on 170- μ m double-side-polished fused silica wafers from Mark Optics. We dice the wafers into 2 cm \times 2 cm chips, and transfer graphene to the chips using 2% poly(methyl methacrylate) (PMMA). We then etch the copper in ferric chloride (Transene, CE-200) for one hour and rinse with five consecutive deionized water baths. We transfer the graphene onto the aluminium-coated chip, and soak overnight in acetone to remove the PMMA. Next, we use photolithography to pattern the pads and evaporate 50 nm of gold. We pattern the graphene strips and etch away the unwanted graphene with a 25-s oxygen plasma. Finally, we soak the chip in a mild (10:1) deionized water/HCl solution until the aluminium release layer has completely disappeared. The chip is transferred directly to a deionized water bath, which is kept refrigerated between uses to discourage bacterial growth.

Atomic force microscope characterization. Atomic force microscope measurements on aluminium-free chips that are run in parallel with measured devices usually give step heights of 1–3 nm above that of pristine exfoliated graphene (Extended Data Fig. 2). Although it is impossible to completely avoid polymer residues from standard transfer and fabrication processing, a 2-nm layer of the stiffest PMMA (Young's modulus $Y = 3.3$ GPa, Poisson ratio $\sigma = 0.4$) should add only about 20 eV to the stiffness (since $\kappa = Yr^3/[12(1-\sigma)]$), which is negligible compared to the measured values.

Influence of surfactants. The addition of a surfactant reduces the graphene's adhesion to the surface and prevents the graphene from permanently sticking to itself. We performed bending stiffness measurements with and without surfactant, and found that the presence of surfactant does not measurably affect the bending stiffness. A surfactant was used in all kirigami experiments. We used sodium dodecylbenzenesulfonate (SDBS) from Sigma-Aldrich (product number 289957), dissolved in deionized water to a concentration of approximately 3 mM. During the measurement process some water evaporates and is replaced with deionized water, so that the concentration of SDBS remains approximately constant over time.

Extraction of $\langle x_{\text{th}}^2 \rangle$ from thermal motion. To extract $\langle x_{\text{th}}^2 \rangle$ from the thermal motion of the gold pad on the free end of the graphene cantilever, we recorded the motion at 90 frames per second for about 20 min to ensure that the entire phase space of the cantilever motion was sampled. The first 20 s from the trace of the free gold pad on a 40 μ m \times 10 μ m cantilever are shown in Extended Data Fig. 3a. We tracked the motion of the pad centroid frame by frame using image analysis to extract the x position of the pad over time; the x direction is perpendicular to the profile of the free gold pad (see inset of Extended Data Fig. 3a). To extract $\langle x_{\text{th}}^2 \rangle$ from this thermal motion, we calculated the power spectral density (PSD), which is the Fourier transform of the autocorrelation of the data, shown in Extended Data

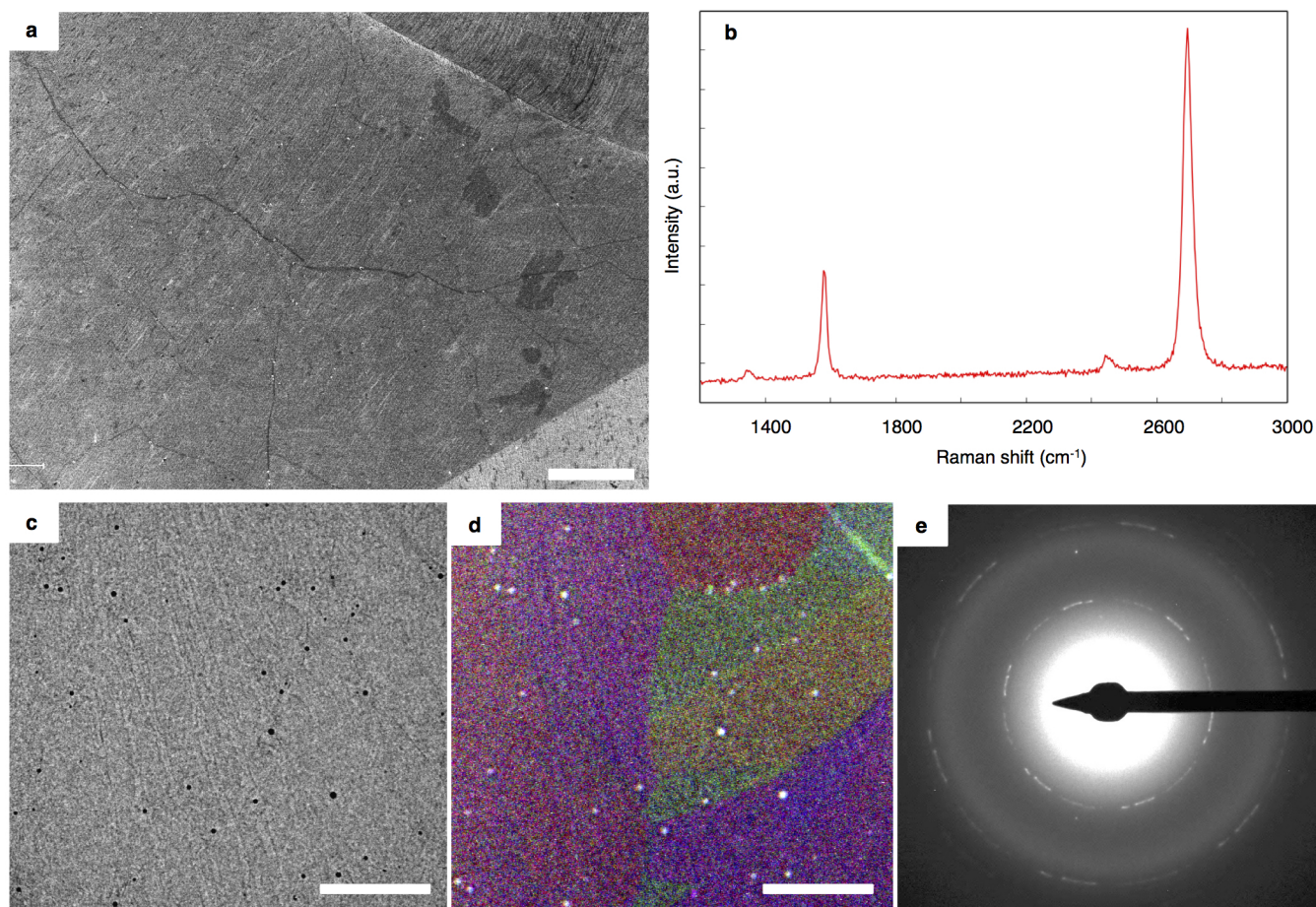
Fig. 3b. In all devices we observed low-frequency $1/f$ noise from the long-timescale motion of the supporting probe (shown in red). This low-frequency noise was excluded from further analysis. We fit the data plotted in blue, which resulted from the thermal motion of the free gold pad, with the theoretical one-sided PSD for Brownian thermal motion^{15,16} (dashed line): $S_{xx}(f) = S_0/[1 + (f/f_c)^2]$, where S_0 is the low-frequency value of the Brownian motion PSD, and f_c is the corner frequency. The integral of this fitted function yields^{15,16} $\langle x_{\text{th}}^2 \rangle: \int_0^\infty S_{xx}(f) df = \langle x_{\text{th}}^2 \rangle$. Using $k = k_B T / \langle x_{\text{th}}^2 \rangle$, where $\langle x_{\text{th}}^2 \rangle = (130 \text{ nm})^2$ for the device shown in Extended Data Fig. 3, we find that the spring constant for this 40 μ m \times 10 μ m cantilever is $k = 2.4 \times 10^{-7} \text{ N m}^{-1}$ and that the bending stiffness $\kappa = 3 \text{ keV}$.

Interferometric measurements. The wavelength of the laser that was used for the interferometric measurements is 436 nm; in water this means that the separation of the black and white bands is $436/4/1.33 = 82 \text{ nm}$. We used a 10-nm full-width-half-maximum bandpass filter with a centre wavelength of 430 nm on the 436-nm line of a mercury arc lamp. The reflectivity of the glass–water–graphene–water cavity creates a situation where the reflectivity changes from 0.0026 to 0.0067 between dark and light bands (based on thin-film equations). A single sheet of graphene in water has a reflectivity of 0.0002, so our geometry greatly enhances the visibility of graphene.

Electrical measurements of stretchable transistors. Electrical measurements were conducted in an approximately 10 mM KCl solution, with a few drops of about 3 mM SDBS solution added. Since water evaporates during the measurement process, we periodically add deionized water to keep the concentration of SDBS approximately constant. The solution was gated by a gold wire, and the gate–drain current was minimized by contacting the drain electrode with a parylene-C-coated tungsten probe. An Ithaco 1211 current preamplifier was used to measure the current, and the system had a negligible gate–drain leakage current of about 10 nA. The device geometry in Fig. 3d is equivalent to approximately 40 squares in series.

Laser force actuation and calibration. Force–displacement curves for cantilevers and a pyramid were measured using the radiation pressure of a 1,064-nm laser. The spring constant was determined by finding the slope of a linear fit to the data. The power of the laser was adjusted using an acousto-optic modulator. The force delivered to a gold pad was calibrated experimentally using the known weight of the gold. For a given power, the laser was focused near the centre of the gold pad, and the change in displacement was measured using a piezo attached to the objective of the optical microscope.

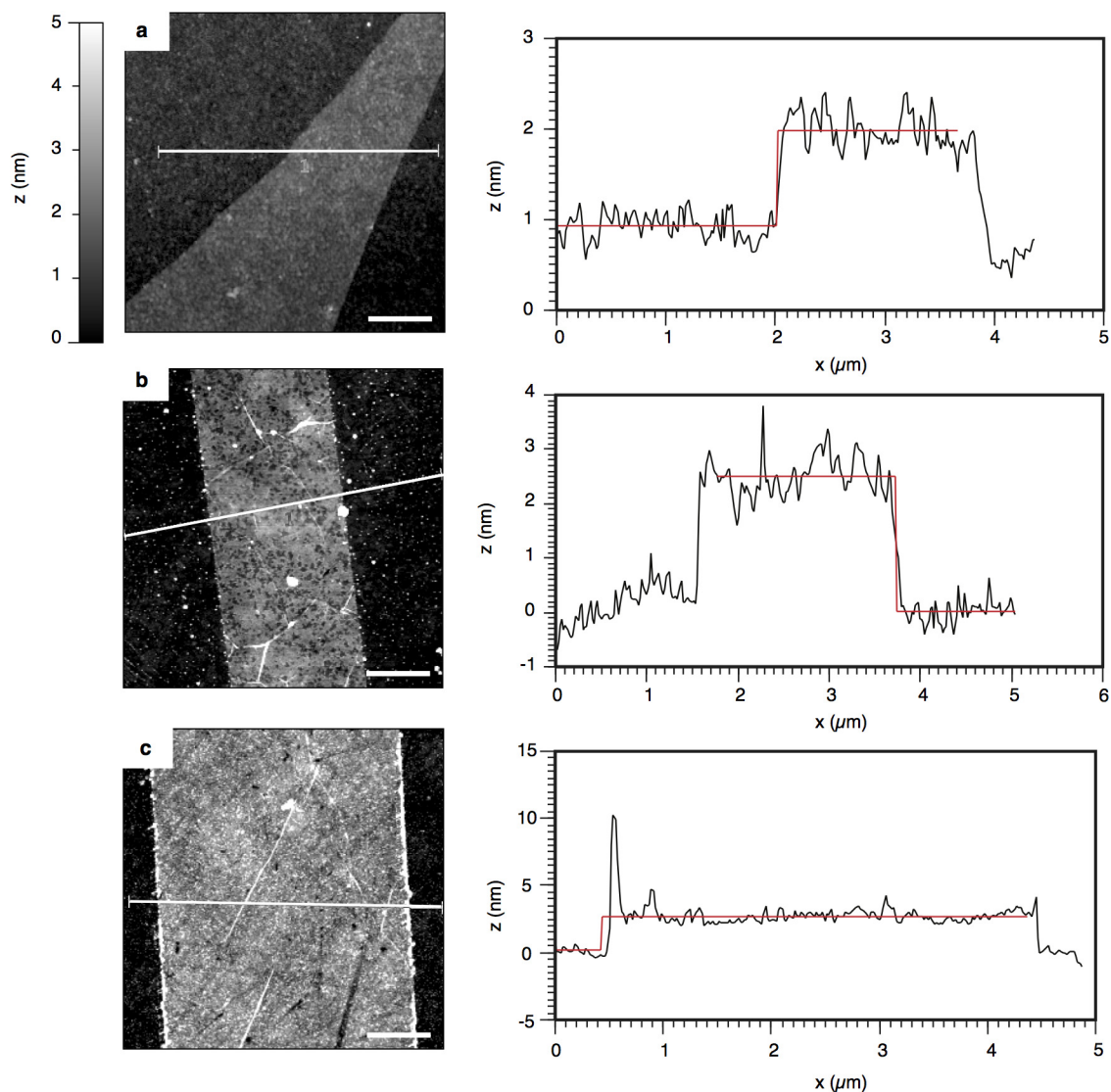
Three-dimensional reconstruction of kirigami devices. We reconstructed the three-dimensional shape of graphene kirigami devices from a z -scanned focal series (Supplementary Video 2). We first acquire a series of images, varying the position of the lens to scan through a z depth of 100 μ m in 10-nm steps. The images are background-subtracted to remove fixed-pattern features on the lens and camera. We then find the focal plane of each x – y pixel by using a refined minimum-intensity algorithm to look for the z plane with the highest contrast. Because the graphene is much thinner than the depth of focus, we assume it behaves like a point object in the z direction. In this model, we use Gaussian fits to find the z centre of the dip in intensity for each x – y pixel. For graphene near the gold pads, we restrict the fit positions to exclude shadowing effects from the pads as they go out of focus. The resulting matrix of z positions is cropped to the size of our object, converted to a three-dimensional matrix of intensities, and smoothed with a three-dimensional Gaussian blur of about 400 nm to reduce noise. Finally, we use TomViz (a development of Paraview that is optimized for tomography visualizations) to render the three-dimensional object. The colour map is based on the intensity of the original video.



Extended Data Figure 1 | Characterization of representative graphene.

a, Scanning electron microscope image of chemical vapour deposition graphene on copper foil. All the graphene used in these experiments was predominantly single-layer, with some small bilayer patches. The larger-scale contrast variations show the copper grains. Scale bar is 10 μm . **b**, Raman spectrum of chemical vapour deposition graphene transferred to SiO_2/Si (285-nm oxide layer) substrate. The spectrum shows graphene's characteristic G peak at $1,580\text{ cm}^{-1}$ and two-dimensional peak at $2,700\text{ cm}^{-1}$; the ratio

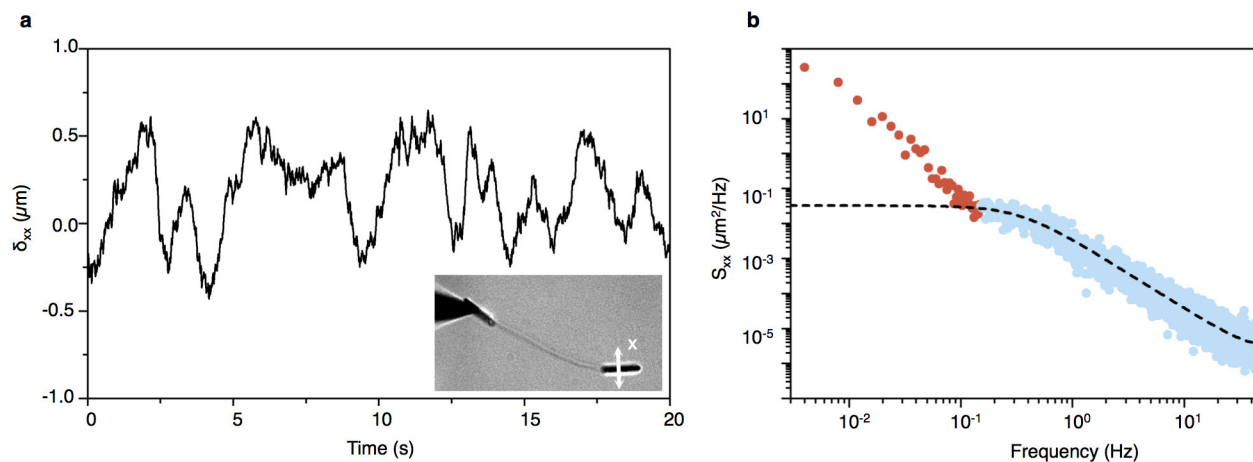
between the two indicates that the graphene is primarily monolayer. A small D peak at $1,350\text{ cm}^{-1}$ indicates low disorder. The small peak at $2,450\text{ cm}^{-1}$ is background. **c**, High-contrast bright-field TEM image of graphene transferred over 10-nm-thick Si_3N_4 windows shows continuous monolayer graphene. Scale bar is 1 μm . **d**, False-colour composite image of dark-field TEM images showing grain size and shape. The graphene is polycrystalline, with grain sizes of the order of micrometres. Scale bar is 1 μm . **e**, TEM diffraction pattern for region shown in **c** and **d**.



Extended Data Figure 2 | Atomic force microscopy of graphene.

a, Exfoliated, unprocessed monolayer graphene. Step height along the red line shown on the height (z) map is 1.0 ± 0.3 nm. **b**, **c**, Representative data from aluminium-free chips that are run in parallel with the devices used in bending stiffness measurements. Step heights are 2.5 ± 0.4 nm and 2.4 ± 0.5 nm.

Chips that look clean under the optical microscope typically have 2–4 nm total step heights. We occasionally see higher residue lines at the edges of the graphene, as in **c**. The PMMA residue is not sufficiently thick to explain the notably increased bending stiffness of the graphene (see Methods). Scale bars are 1 μ m.

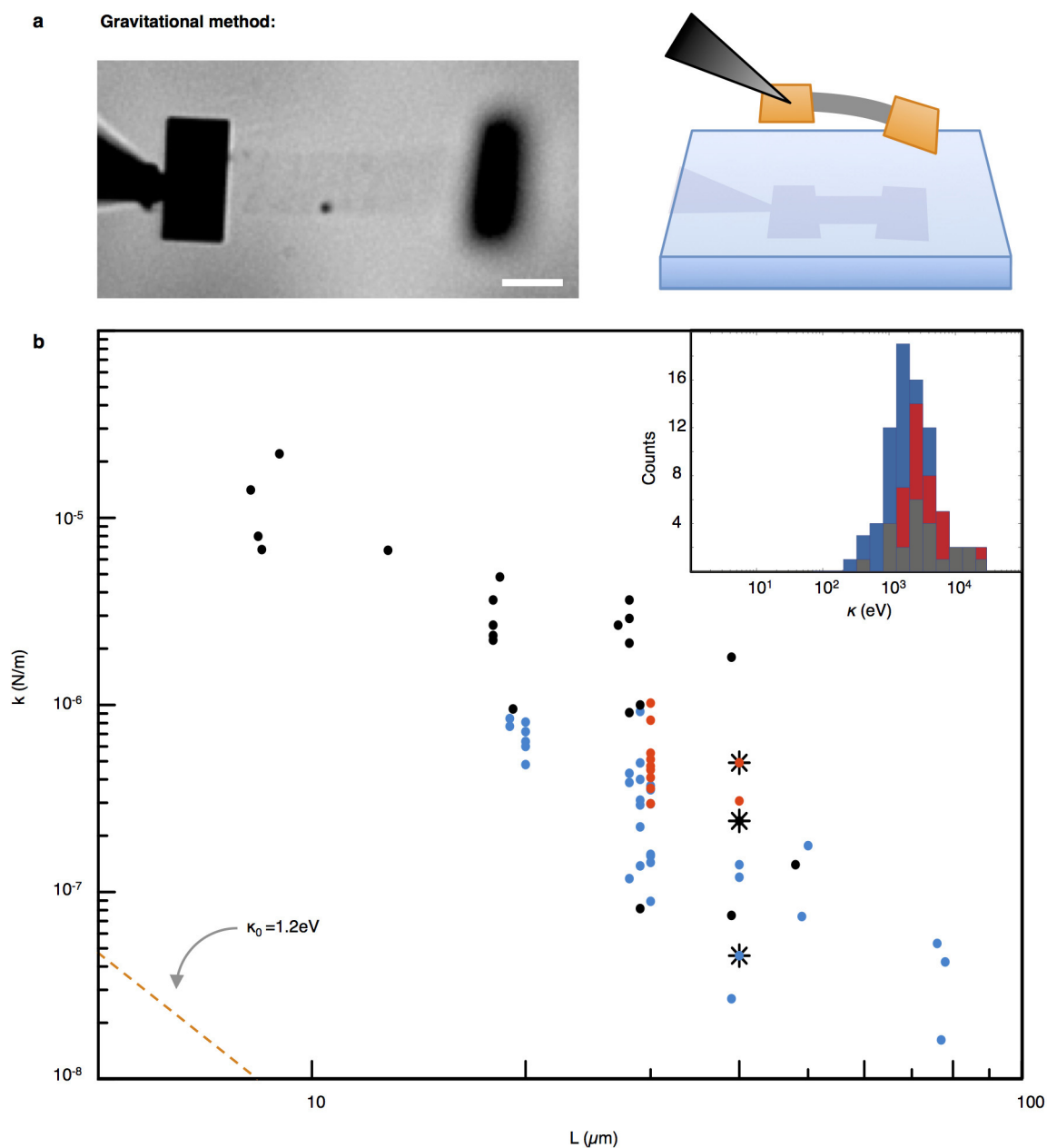


Extended Data Figure 3 | Thermal motion of graphene cantilever gold pads.

a, Time trace of the x position of the gold pad centroid on a $40\ \mu\text{m} \times 10\ \mu\text{m}$ graphene cantilever, showing the first 20 s of a 20-min trace. The x direction is perpendicular to the profile of the gold pad, as indicated in the inset.

b, PSD (S_{xx}) of the full 20-min time trace from **a**. The blue data points were included in the fit to the Brownian motion PSD function (dashed line); the red

data points from about 10^{-3} Hz to about 10^{-1} Hz were excluded, because they show considerable $1/f$ noise as a result of the motion of the probe holding the cantilever. We integrate the fitted function S_{xx} to determine $\langle x_{\text{th}}^2 \rangle$. For the device shown, $\langle x_{\text{th}}^2 \rangle = (130\ \text{nm})^2$, the spring constant $k = 2.4 \times 10^{-7}\ \text{N m}^{-1}$, and the bending stiffness $\kappa = 3\ \text{keV}$.



Extended Data Figure 4 | Bending stiffness measurements. **a**, We also performed a rough measurement of the spring constant using the force of gravity on the gold pads. After lifting the cantilever off the surface, the vertical deflection x_g is determined using the shallow depth of focus of the microscope, adjusted for the change in index of refraction. The gravitational force F_g (corrected for buoyancy) yields the spring constant: $k = F_g/x_g$. For the 50- μm -long cantilever shown (scale bar is 10 μm) with a 2-pN gold pad and $x_g = 25 \mu\text{m}$, we find that $k = 8 \times 10^{-8} \text{ N m}^{-1}$. We repeated the measurement for a variety of devices of varying length L and width $W = 10 \mu\text{m}$. We have

observed that the cantilevers sometimes curve downwards even in the absence of an applied force, presumably due to residual materials or strains in the graphene, so these gravitational measurements probably have a systematic offset. **b**, The measured spring constants of 10- μm -wide devices are shown on a plot of spring constant versus device length for the thermal fluctuation (black), gravitational deflection (blue), and laser force measurements (red). The data from all three techniques (plus additional laser data for devices with other widths, as in Fig. 2) are shown in the inset as a histogram. Data with stars are from the same device, using the three different measurement techniques.

Dosage delivery of sensitive reagents enables glove-box-free synthesis

Aaron C. Sather¹, Hong Geun Lee¹, James R. Colombe¹, Anni Zhang¹ & Stephen L. Buchwald¹

Contemporary organic chemists employ a broad range of catalytic and stoichiometric methods to construct molecules for applications in the material sciences¹, and as pharmaceuticals^{2–5}, agrochemicals, and sensors⁶. The utility of a synthetic method may be greatly reduced if it relies on a glove box to enable the use of air- and moisture-sensitive reagents or catalysts. Furthermore, many synthetic chemistry laboratories have numerous containers of partially used reagents that have been spoiled by exposure to the ambient atmosphere. This is exceptionally wasteful from both an environmental and a cost perspective. Here we report an encapsulation method for stabilizing and storing air- and moisture-sensitive compounds. We demonstrate this approach in three contexts, by describing single-use capsules that contain all of the reagents (catalysts, ligands, and bases) necessary for the glove-box-free palladium-catalysed carbon–fluorine^{7–9}, carbon–nitrogen^{10,11}, and carbon–carbon¹² bond-forming reactions. This strategy should reduce the number of error-prone, tedious and time-consuming weighing procedures required for such syntheses and should be applicable to a wide range of reagents, catalysts, and substrate combinations.

We sought to develop a system to allow for the bench-top storage of pre-measured quantities of air- and moisture-sensitive reagents and catalysts in such a way that the contained material would be liberated into a reaction mixture upon subjection to typical reaction conditions. We initially chose paraffin wax as a stabilizing agent as it has been shown to be an effective material for protecting sensitive compounds from oxygen and water in the atmosphere^{13–16}. For instance, a paraffin wax dispersion of normally pyrophoric potassium hydride can be easily handled and is relatively stable under ambient laboratory conditions¹⁷. As such, preliminary work focused on creating dispersions of reagent and reagent mixtures using molten paraffin wax, although it was not possible to achieve a uniform distribution of the components using this method. Upon cooling, a gradient was established within the paraffin matrix, making it impossible to determine the concentration of the constituents for a given sample. Moreover, reagents located on the surface are exposed to the atmosphere, and free to react with air and water. To address these shortcomings, we developed a simple method to enclose premeasured amounts of catalysts and reagents within paraffin capsules, isolating them from the atmosphere. Hollow paraffin (melting point 58–62 °C) shells were manually prepared and filled with catalyst and reagent combinations, thus providing a single stabilized entity with which to conveniently carry out a variety of transformations (Supplementary Figs 1–4).

To probe the effectiveness of the encapsulation technology, we first studied the oxygen- and moisture-sensitive palladium-catalysed nucleophilic fluorination of aryl triflates (ArOTf) (Fig. 1a)^{7,8}. Fluorinated aromatics are a common motif found in pharmaceuticals and agrochemicals, and are introduced to impart metabolic stability and enhanced lipophilicity¹⁸. The introduction of a fluorine atom can also increase protein-binding affinity¹⁹ and affect the orientation and conformation of a molecule when binding to a protein²⁰. As a result, the synthesis of fluorinated compounds has generated great interest²¹.

Traditional methods^{22,23} of incorporating a fluorine atom onto an aromatic ring typically require harsh conditions, which limits the scope of these transformations and necessitates the introduction of fluorine early in the synthesis. In contrast, palladium catalysis allows for the late-stage transformation of ArOTf and aryl bromides (ArBr) to the corresponding aryl fluoride (Ar–F), providing good yields and exhibiting a much broader substrate scope. In addition to the well-documented challenges associated with this transformation^{24,25}, which includes a difficult reductive elimination (RE) step, care must be taken to exclude water to prevent proto-demetalation (ArH) and formation of phenol (ArOH) and biaryl ether (Ar₂O) side products. The metal fluoride salts (caesium fluoride (CsF) and silver(I) fluoride (AgF)) used in these reactions are hygroscopic, and the Pd(0) precatalyst is sensitive towards oxygen⁸, which requires the reaction to be set up in a glove box.

To address problems arising from stability, the hollow paraffin shells were charged with 2 mol% **P1** (4 mol% of Pd) using **L1** as the supporting ligand and 3 mmol of CsF (Fig. 1b, blue capsule), and stored on the bench top. With the capsules in hand, the reaction set-up is inherently simple. The desired ArOTf (1 mmol) is added to an oven-dried reaction tube equipped with a stir bar, followed by a capsule. After evacuating the tube and backfilling with argon, solvent is added. Upon heating to the specified temperature, the capsule melts and releases its contents, initiating the transformation. When the reaction is complete, the paraffin is easily removed by precipitation, filtration, and silica gel chromatography.

With this method, a variety of aryl (1–4) and heteroaryl (5, 6) fluorides could be prepared in yields that are comparable to those obtained with the aid of a glove box (Fig. 1c). While some examples were previously reported using lower catalyst loadings (2–3 mol% Pd), 2 mol% **P1** (4 mol% Pd) was loaded into each capsule to provide a universal reagent capable of transforming all desired ArOTfs—facilitating operational simplicity. To demonstrate the robustness of this technology, a capsule was suspended in a beaker of water for 24 h, dried with a paper towel, and used in a reaction to provide the Ar–F in undiminished yield (Fig. 1c, 3). However, a capsule that was kept on the bench top at room temperature for over eight months showed decreased activity and required elevated reaction temperatures to achieve full conversion of the starting material (Supplementary Table 1).

With this initial success, we applied the capsule method to the palladium-catalysed nucleophilic fluorination of aryl bromides (ArBr) (Fig. 2). As previously described⁹, two fluoride salts are required for this transformation (KF and AgF), as well as a palladium(0) precatalyst with either **L2** or **L1** as the supporting ligand. Because the **L2**-supported precatalyst (**P2**) is effective for the fluorination of both aryl and heteroaryl bromides, it was selected as the optimal catalyst for use with the wax capsules. As in the preceding example, the hollow paraffin shells were charged with both **P2** and the reagents necessary to transform 1 mmol of ArBr to the desired Ar–F (Fig. 2a, red capsule).

These three-component capsules were able to provide a range of aryl (7–9) and heteroaryl (10–12) fluorides from commercially

¹Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

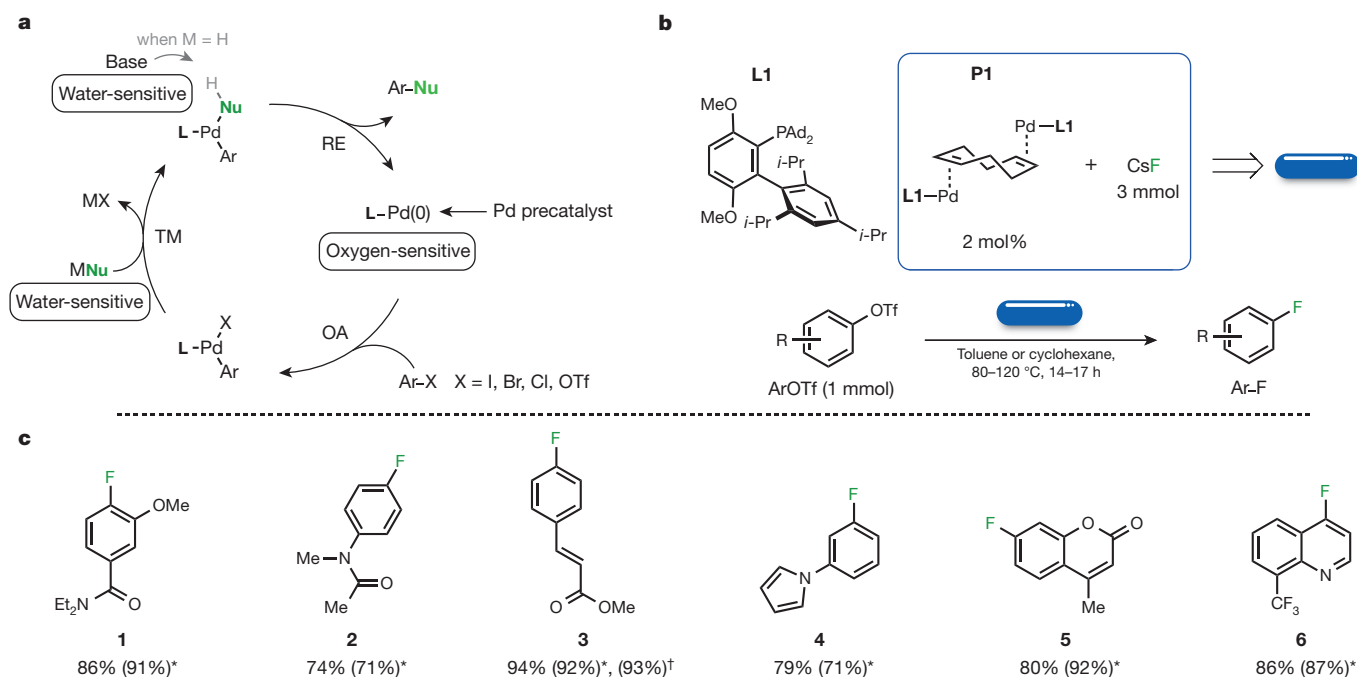


Figure 1 | Wax capsules for the glove-box-free Pd-catalysed nucleophilic fluorination of aryl triflates. **a**, The catalytic cycle of a typical palladium-catalysed cross-coupling reaction; sensitive aspects are highlighted for clarity. **L** is a ligand, OA is oxidative addition, TM is transmetalation, and RE is reductive elimination. **M** is either a counter cation or a proton. COD is 1,5-cyclooctadiene. **b**, Contents of the wax capsule for the fluorination of

ArOTf. The green F (fluorine) highlights the site of the transformation.

c, Glove-box-free fluorination of ArOTf. Isolated yields are reported as an average of two runs. *Isolated yields that were previously reported and obtained using a glove box to set up the reactions. †Isolated yield after soaking a capsule in water for 24 h.

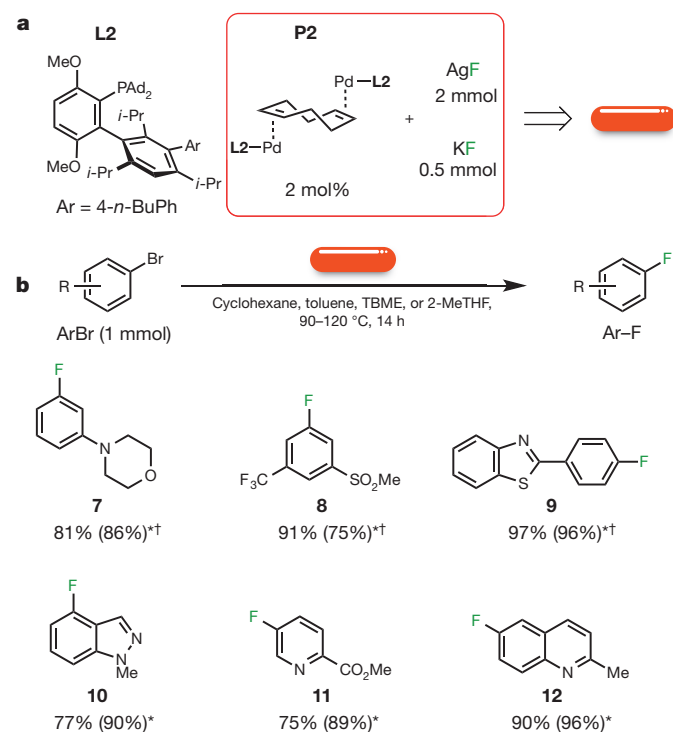


Figure 2 | Wax capsules for the glove-box-free Pd-catalysed nucleophilic fluorination of aryl bromides. Ad is 1-adamantyl. THF, tetrahydrofuran. TBME, *tert*-butyl methyl ether. The green F (fluorine) highlights the site of the transformation. **a**, Contents of the wax capsules for the fluorination of aryl bromides. **b**, Glove-box-free fluorination of aryl bromides. Isolated yields are reported as an average of two runs. *Isolated yields that were previously reported and obtained using a glove box to set up the reactions †Previously reported isolated yields obtained using **L1** as the supporting ligand.

available ArBr in good yields, which rival those that were obtained when the reactions were set up in a glove box (Fig. 2b). Again, to test the capsules' robustness, a capsule was placed in a beaker of water for 24 h. Once dried, the activity of this capsule matched that of a capsule that never made direct contact with water (Supplementary Table 3).

To highlight the generality of this approach, we applied the paraffin capsule technology to other reaction types that are useful in a variety of research areas. The first method we pursued was the palladium-catalysed cross-coupling of aryl halides with amine nucleophiles²⁶, which has become an indispensable tool for applications in materials science¹, sensor synthesis⁶, and pharmaceutical development²⁻⁵. Over the years, our laboratory has developed a series of biaryl monophosphine ligands and highly efficient base-activated, ligated Pd(II) precatalysts for C-N bond formation that are commercially available¹⁰.

Although the components of this reaction are not sensitive to oxygen, the base required is hygroscopic, and must be kept in a glove box or stored in a desiccator. Additionally, it was discovered that a dual ligand mixture composed of **L3** and **L4** yielded a system capable of coupling both primary and secondary amine nucleophiles¹¹. Thus, a paraffin capsule containing **L3**-based precatalyst (**P3**), **L4**, and sodium *tert*-butoxide (base, Fig. 1a) would be capable of coupling a breadth of primary and secondary amines by the addition of a single universal encapsulated reagent (Fig. 3a, orange capsule), eliminating the need for time-consuming reaction optimizations. Indeed, these capsules coupled a primary alkyl amine (**13**), an acyclic secondary amine (**14**), a cyclic secondary amine (**15**), anilines (**16** and **17**), and a primary hetero-aromatic amine (**18**) to aryl halides and heteroaryl chlorides (Fig. 3b). The capsules were stored on the bench top and showed no signs of degradation over a period of over eight months, even though the base-activated **P3** was stored in close contact with sodium *tert*-butoxide (Fig. 3b, **18**).

The palladium-catalysed Negishi cross-coupling of 2-pyridylzinc dioxanate with aryl halides and triflates was also adapted for use with

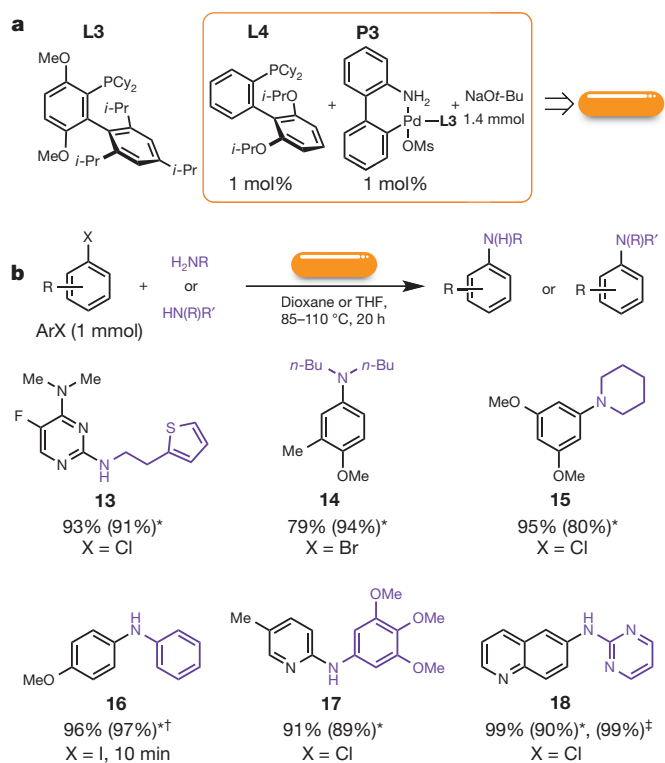


Figure 3 | Wax capsules for the Pd-catalysed CN cross-coupling of 1° and 2° amines with aryl halides. Ms, methanesulfonate. **a**, Contents of the wax capsule for the amination of aryl halides. **b**, Examples of C–N coupling using the wax capsules. The purple portion of the molecule indicates the amine coupling partner, highlighting the site of the transformation. Isolated yields are reported as an average of two runs. *Previously reported isolated yields. †Toluene was used as the reaction solvent. ‡Isolated yield after storing a capsule on the bench top for over eight months.

paraffin wax capsules¹². The 2-pyridyl group has found applications in functional materials²⁷ and is a component of biologically active compounds²⁸. Traditional 2-pyridyl nucleophiles such as boronates suffer from instability²⁹, which makes them difficult to employ in Suzuki–Miyaura cross-coupling reactions. In contrast, the dioxane-stabilized 2-pyridylzinc reagent is a solid, competent nucleophile that can be briefly manipulated in air, although prolonged storage is problematic owing to its sensitivity to water¹².

Encapsulation of the basic 2-pyridylzinc dioxanate (MNu, Fig. 1a) with base-activated palladium precatalyst (**P4**) within a paraffin wax capsule provides a bench-top-stable reagent and an efficient means of introducing this important functional group to a variety of (hetero)-aryl halides and triflates (Fig. 4a, purple capsule). With this technology, (hetero)aryl chlorides (**19** and **20**), aryl triflate (**21**), and (hetero)aryl bromides (**22**, **23**, and **24**) were easily converted to the desired 2-pyridyl compounds. To demonstrate the stability of the zinc reagent, capsules containing 2-pyridylzinc dioxanate that have been stored on the bench top for one year were shown by titration to contain the original amount of active material (Supplementary Table 5).

We have reported that several valuable oxygen- and water-sensitive cross-coupling catalysts and reagents can be stabilized by encapsulation within inert, hydrophobic wax capsules. These capsules provide access to an array of desirable cross-coupled products by the convenient addition of a single, user-friendly, bench-top-stable reagent. Through collaboration with chemical providers, the manual capsule preparation process should be easy to mechanize for large-scale production, making this technology widely available for a variety of traditionally sensitive compounds³⁰. Furthermore, we envision that this concept will transform other moisture- and air-sensitive reagents

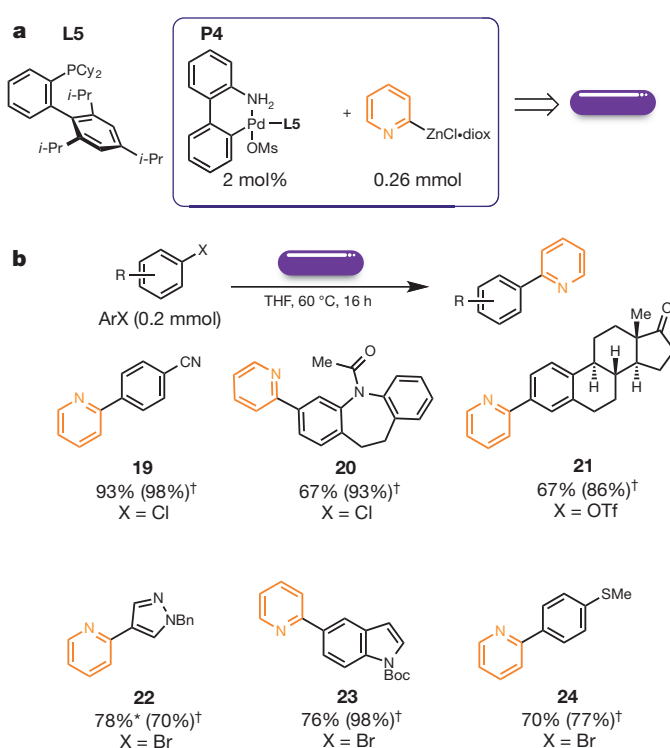


Figure 4 | Wax capsules for the Pd-catalysed Negishi cross-coupling of 2-pyridylzinc dioxanate. Diox, dioxane. Boc, *tert*-butoxycarbonyl. The orange portion of the molecule indicates the pyridine coupling partner, highlighting the site of the transformation. **a**, Contents of the wax capsule for the Negishi cross-coupling of 2-pyridylzinc. **b**, Examples of Negishi cross-coupling using the wax capsules. Isolated yields are reported as an average of two runs. *As previously reported, an additional 4 mol% **L5** and 2 mol% **P4** were added. †Previously reported isolated yields.

(such as ZnCl_2 , AlCl_3 , AgF_2) by turning reactions that employ these into operationally simpler and more robust processes.

Received 27 April; accepted 10 June 2015.

- Ozin, G. A., Arsenault, A. C. & Cademartiri, L. (eds) *Nanochemistry: A Chemical Approach to Nanomaterials* 2nd edn (Royal Society of Chemistry, 2009).
- Buitrago Santanilla, A. *et al.* Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science* **347**, 49–53 (2015).
- Affouard, C. *et al.* Multi-kilo delivery of AMG 925 featuring a Buchwald–Hartwig amination and processing with insoluble synthetic intermediates. *Org. Process Res. Dev.* **19**, 476–485 (2015).
- Magano, J. & Dunetz, J. R. Large-scale applications of transition metal-catalyzed couplings for the synthesis of pharmaceuticals. *Chem. Rev.* **111**, 2177–2250 (2011).
- Crawley, M. L. & Trost, B. M. (eds) *Applications of Transition Metal Catalysis in Drug Discovery and Development* (Wiley, 2012).
- Wang, B. & Anslyn, E. V. (eds) *Chemosensors Principles, Strategies, and Applications* (Wiley, 2011).
- Watson, D. A. *et al.* Formation of ArF from LPdAr(F): catalytic conversion of aryl triflates to aryl fluorides. *Science* **325**, 1661–1664 (2009).
- Lee, H. G., Milner, P. J. & Buchwald, S. L. An improved catalyst system for the Pd-catalyzed fluorination of (hetero)aryl triflates. *Org. Lett.* **15**, 5602–5605 (2013).
- Lee, H. G., Milner, P. J. & Buchwald, S. L. Pd-catalyzed nucleophilic fluorination of aryl bromides. *J. Am. Chem. Soc.* **136**, 3792–3795 (2014).
- Bruno, N. C., Tudge, M. T. & Buchwald, S. L. Design and preparation of new palladium precatalysts for C–C and C–N cross-coupling reactions. *Chem. Sci.* **4**, 916–920 (2013).
- Fors, B. P. & Buchwald, S. L. A multiligand based Pd catalyst for C–N cross-coupling reactions. *J. Am. Chem. Soc.* **132**, 15914–15917 (2010).
- Colombe, J. R., Bernhardt, S., Stathakis, C., Buchwald, S. L. & Knochel, P. Synthesis of solid 2-pyridylzinc reagents and their application in Negishi reactions. *Org. Lett.* **15**, 5754–5757 (2013).
- Kosak, K. M. & Kosak, M. K. Preparation of wax beads containing a reagent for release by heating. US patent 5,413,924 (1995).

14. Taber, D. F. & Li, H.-Y. Novel stable compositions of water and oxygen sensitive compounds and their method of preparation. US patent publ. no. 2005/0288257 A1 (2005).
15. Fang, Y. *et al.* A new chromium-based catalyst coated with paraffin for ethylene oligomerization and the effect of chromium state on oligomerization selectivity. *Appl. Catal. A* **235**, 33–38 (2002).
16. Taber, D. F. & Frankowski, K. J. Grubbs' catalyst in paraffin: an air-stable preparation for alkene metathesis. *J. Org. Chem.* **68**, 6047–6048 (2003).
17. Taber, D. F. & Nelson, C. G. Potassium hydride in paraffin: a useful base for organic synthesis. *J. Org. Chem.* **71**, 8973–8974 (2006).
18. Smart, B. E. Fluorine substituent effects (on bioactivity). *J. Fluor. Chem.* **109**, 3–11 (2001).
19. Paulini, R., Müller, K. & Diederich, F. Orthogonal multipolar interactions in structural chemistry and biology. *Angew. Chem. Int. Ed.* **44**, 1788–1805 (2005).
20. Müller, K., Faeh, C. & Diederich, F. Fluorine in pharmaceuticals: looking beyond intuition. *Science* **317**, 1881–1886 (2007).
21. Campbell, M. G. & Ritter, T. Modern carbon-fluorine bond forming reactions for aryl fluoride synthesis. *Chem. Rev.* **115**, 612–633 (2015).
22. Balz, G. & Schiemann, G. Aromatic fluorine compounds. I. A new method for their preparation. *Ber. Dtsch. Chem. Gesell.* **60**, 1186–1190 (1927).
23. Finger, G. C. & Kruse, C. W. Aromatic fluorine compounds. VII. Replacement of aromatic -Cl and -NO₂ groups by -F. *J. Am. Chem. Soc.* **78**, 6034–6037 (1956).
24. Grushin, V. V. The organometallic fluorine chemistry of palladium and rhodium: studies toward aromatic fluorination. *Acc. Chem. Res.* **43**, 160–171 (2010).
25. Yandulov, D. V. & Tran, N. T. Aryl-fluoride reductive elimination from Pd(II): feasibility assessment from theory and experiment. *J. Am. Chem. Soc.* **129**, 1342–1358 (2007).
26. De Meijere, A. & Diederich, F. (eds) *Metal-Catalyzed Cross-Coupling Reactions* 2nd edn (Wiley, 2004).
27. Wild, A., Winter, A., Schlütter, F. & Schubert, U. S. Advances in the field of π -conjugated 2,2':6',2"-terpyridines. *Chem. Soc. Rev.* **40**, 1459–1511 (2011).
28. DeGoey, D. A. *et al.* 2-Pyridyl P1'-substituted symmetry-based human immunodeficiency virus protease inhibitors (A-792611 and A-790742) with potential for convenient dosing and reduced side effects. *J. Med. Chem.* **52**, 2571–2586 (2009).
29. Lennox, A. J. J. & Lloyd-Jones, G. C. The slow-release strategy in Suzuki–Miyaura coupling. *Isr. J. Chem.* **50**, 664–674 (2010).
30. Michaudel, Q., Ishihara, Y. & Baran, P. S. Academia-industry symbiosis in organic chemistry. *Acc. Chem. Res.* **48**, 712–721 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements Research reported in this publication was supported by the National Institutes of Health under award number R01GM46059. A.C.S. thanks the National Institutes of Health for a postdoctoral fellowship (1F32GM108092-01A1). J.R.C. thanks the National Science Foundation for a pre-doctoral fellowship (1122374). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank Y. Ye for providing samples of **L2**. We also thank M. Pirnot, Y. Wang, and C. Nguyen for assistance with the preparation of the manuscript.

Author Contributions S.L.B. had the idea to encapsulate reagents in wax; A.C.S. and H.G.L. invented the capsules and designed their preparation; A.C.S., H.G.L., and S.L.B. designed the research; A.C.S., H.G.L., and J.R.C. prepared the filled capsules; A.C.S., H.G.L., J.R.C., and A.Z. performed the experiments; A.C.S. wrote the manuscript. All authors commented on the final draft of the manuscript and contributed to the analysis and interpretation of the data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper in the HTML version. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.L.B. (sbuchwal@mit.edu).

The role of ridges in the formation and longevity of flat slabs

Sanja Knezevic Antonijevic¹, Lara S. Wagner², Abhash Kumar¹, Susan L. Beck³, Maureen D. Long⁴, George Zandt³, Hernando Tavera⁵ & Cristobal Condori⁵

Flat-slab subduction occurs when the descending plate becomes horizontal at some depth before resuming its descent into the mantle. It is often proposed as a mechanism for the uplifting of deep crustal rocks ('thick-skinned' deformation) far from plate boundaries, and for causing unusual patterns of volcanism, as far back as the Proterozoic eon¹. For example, the formation of the expansive Rocky Mountains and the subsequent voluminous volcanism across much of the western USA has been attributed to a broad region of flat-slab subduction beneath North America that occurred during the Laramide orogeny (80–55 million years ago)². Here we study the largest modern flat slab, located in Peru, to better understand the processes controlling the formation and extent of flat slabs. We present new data that indicate that the subducting Nazca Ridge is necessary for the development and continued support of the horizontal plate at a depth of about 90 kilometres. By combining constraints from Rayleigh wave phase velocities with improved earthquake locations, we find that the flat slab is shallowest along the ridge, while to the northwest of the ridge, the slab is sagging, tearing, and re-initiating normal subduction. On the basis of our observations, we propose a conceptual model for the temporal evolution of the Peruvian flat slab in which the flat slab forms because of the combined effects of trench retreat along the Peruvian plate boundary, suction, and ridge subduction. We find that while the ridge is necessary but not sufficient for the formation of the flat slab, its removal is sufficient for the flat slab to fail. This provides new constraints on our understanding of the processes controlling the beginning and end of the Laramide orogeny and other putative episodes of flat-slab subduction.

Oceanic plates subduct at different angles ranging from steep to shallow, with flat slabs representing the horizontal endmember. The subduction of buoyant aseismic ridges and plateaus comprising over-thickened oceanic crust has long been thought to play a part in the formation of flat slabs³. More recent work has identified other potential contributing factors, including trench retreat^{4,5}, rapid overriding plate motion^{4,5}, and suction between the flat slab and overriding continental mantle lithosphere⁵. Many of these studies do not preclude the need for additional buoyancy from overthickened oceanic crust. However, a few recent studies suggest that subducting ridges do not affect the formation or sustainability of flat slabs^{6,7}.

To evaluate the influence of subducting ridges on the evolution of flat slabs, we focus on the flat slab in southern Peru (Fig. 1). Here, the subducting Nazca Ridge trends at an oblique angle to relative plate motion, resulting in a northward migration of the overriding continent relative to the down-going ridge⁸. We have collected and analysed data from two deployments of broadband seismometers in central and southern Peru: PULSE (Peru Lithosphere and Slab Experiment)⁹, and CAUGHT (Central Andean Uplift and Geodynamics of High Topography)¹⁰. We also incorporate data from eight stations from the PERUSE deployment (Peru Slab Experiment)¹¹ and the permanent

station called NNA in Lima, Peru (Fig. 1). Here we present a three-dimensional model of shear-wave velocity structure between -10° and -18° , obtained from the inversion of earthquake-generated Rayleigh-wave-phase velocities (Fig. 2 and Extended Data Figs 2–10). We also relocate slab seismicity across our study area using a double difference methodology (Figs 2 and 3, Extended Data Fig. 1 and Supplementary Table 1) (see Methods for details).

Our tomographic images and improved earthquake locations show the flat slab to be shallowest along the present-day projected location of the subducted Nazca Ridge (Figs 2g and 3g, h). To the south (Fig. 2h), the slab transitions abruptly from flat to normal, and earthquake locations align with an increase in shear-wave velocity in our model. To the north, where previous studies have proposed a broad flat slab of relatively uniform depth^{12,13}, we see a gradual but marked deepening of the plane of seismicity associated with subducted slab known as the Wadati–Benioff zone (Figs 2e, f and 3g, h). To the east, high shear-wave velocities associated with the flat slab extend substantially inboard (that is, away from the trench, inland) than the seismically active portion of the plate (Figs 2g and 3g, h). The downward bend

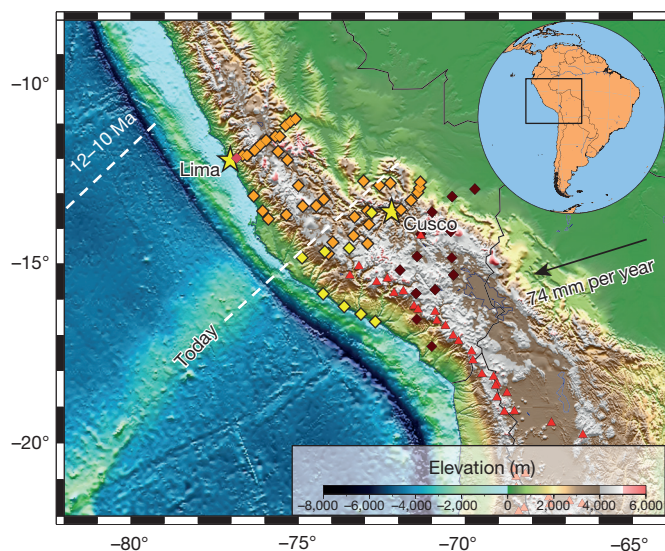


Figure 1 | Reference map of the Peruvian flat-slab region, illustrating the subducting Nazca Ridge beneath the advancing South American plate.

Diamonds represent seismic stations used in this study: orange, PULSE; dark red, CAUGHT; yellow, PERUSE; red, the permanent NNA station. Yellow stars represent the cities Lima and Cusco. Red triangles represent volcanoes active during the Holocene epoch. The black arrow indicates the relative motion of the South American plate with respect to the Nazca Plate²¹. Dotted white lines show the estimated position of the Nazca Ridge 12–10 Ma and today⁸.

¹Department of Geological Sciences, University of North Carolina at Chapel Hill, CB 3315, Chapel Hill, North Carolina 27599, USA. ²Department of Terrestrial Magnetism, Carnegie Institution for Science, 5241 Broad Branch Road NW, Washington DC 20015, USA. ³Department of Geosciences, University of Arizona, 1040 East 4th Street, Tucson, Arizona 85721, USA. ⁴Department of Geology and Geophysics, Yale University, 210 Whitney Avenue, New Haven, Connecticut 06511, USA. ⁵Instituto Geofísico del Perú, Calle Badajoz 169, Lima 15012, Peru.

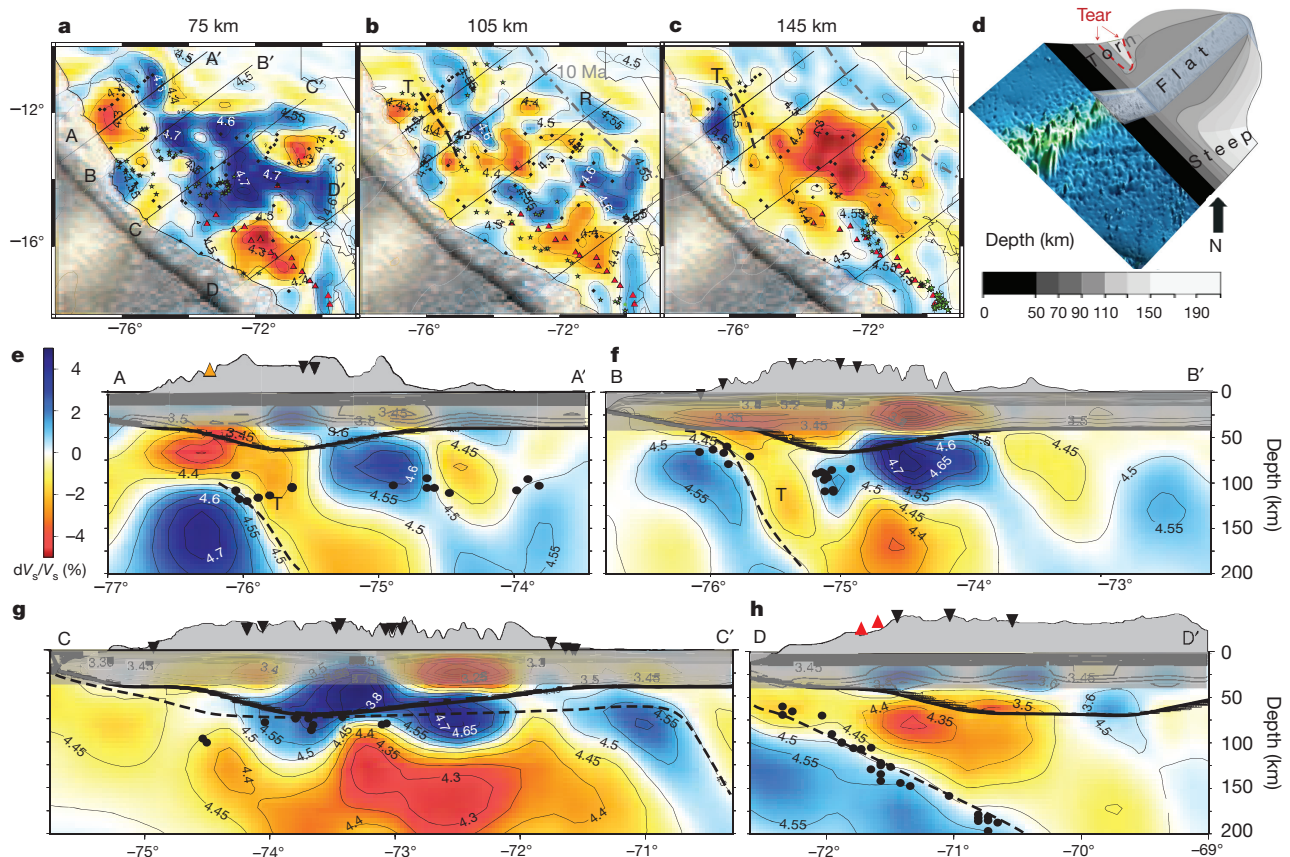


Figure 2 | Three-dimensional model of the structure of shear-wave velocities between -10° and -18° . **a–c**, Shear-wave velocities and seismicity at depths of 75 km (**a**), 105 km (**b**) and 145 km (**c**), and transects along the northern reinitiating steep slab (A–A', B–B'), flat slab (C–C') and southern steep slab (D–D') segments. Colours indicate velocity deviations, dV_s/V_s (%); contours show absolute velocities in kilometres per second (numbered). **a–c**, Black circles represent stations used in our study; red triangles are Holocene volcanoes; green stars are earthquakes within 20 km of the depth shown; black lines refer to cross-sections shown in **e–h**. The grey dashed line in **b** and **c** shows the location of the trench 10 Ma (ref. 8);

the black dashed line (labelled 'T') indicates the location of the slab tear. 'R' refers to the resumption of steep subduction at the eastern edge of the flat slab. **d**, Inferred flat-slab geometry along the Nazca Ridge track, and slab tear north of the ridge. **e–h**, Cross-sections of slab segments shown in **a–c**. Black dots show earthquake locations from this study; black inverted triangles are stations; red triangles are Holocene volcanoes; orange triangle represents the location of a measurement of unusually high heat flow¹⁵. Dashed lines show the inferred top of the slab. The thick black line shows the crustal thickness.

in the high-velocity plate at the easternmost extent of the flat slab appears to coincide with the location of the Peruvian trench about 10 million years ago (Ma)⁸.

Of particular note is the geometry of the subducted plate north of the projected Nazca Ridge track (Fig. 2a–c, e, f). Here, we observe a dipping high-velocity anomaly to the trenchward side of a dipping low-velocity anomaly. We note the similarity between these structures (in an area previously believed to comprise typical flat slab) and those observed to the south beneath the active arc (Fig. 2e, f, h). We also note the difference between these structures and those adjacent to the ridge, where the continuous flat slab is well resolved (Fig. 2e–g and Extended Data Figs 3, 6–10). We interpret the westward-dipping low-velocity region parallel to the trench to be evidence of asthenosphere (the viscous, weak region of the upper mantle) between two torn portions of subducted plate. The dipping high-velocity anomaly to the west indicates the presence of a normally dipping slab extending to a depth of at least 200 km. This is consistent with the location of shear-wave scatterers identified from converted phases in earlier studies¹⁴. We propose that the subhorizontal seismicity to the east of the tear is located in remnant flat slab that has not yet been fully subducted. Local shear wave splitting studies show that shear waves move faster parallel to the trench⁹, consistent with north–south-directed asthenospheric flow through a break in the Nazca plate. We also note the

presence of a localized high heat flow (196 mW s^{-2}) above this low-velocity anomaly¹⁵ (Fig. 2e). Along the northernmost transect, the location of the slab is not well resolved above a depth of about 100 km (Fig. 2e). Future work using ambient noise tomography may help us to resolve the slab geometry here, by providing improved constraints on velocities at shallower depths.

We incorporate the results of previous geodynamic modelling studies with our results, to create a conceptual model of the temporal evolution of the Peruvian flat slab (Fig. 3). We begin with the initiation of ridge subduction at approximately 11.2 Ma (ref. 8), before which we assume normal subduction across our study area (Fig. 3a). From there, we base our proposed temporal evolution of the Peruvian flat slab on four principles.

First, we present our conceptual model from the reference frame of a laterally stationary Nazca plate. Second, while most of the Nazca plate sinks vertically at a relatively constant rate, the plate containing the Nazca Ridge ceases to sink at a depth of about 90 km (Fig. 3b–f). We propose that this is due to buoyancy imparted by the overthickened oceanic crust and harzburgite layer associated with the ridge, consistent with previous modelling studies¹⁶. Third, we observe that the modern inboard extent of the Peruvian flat slab corresponds to the location of the trench at about 10 Ma. Given that the projected location of the Nazca Ridge extends further to the east, this finding suggests that

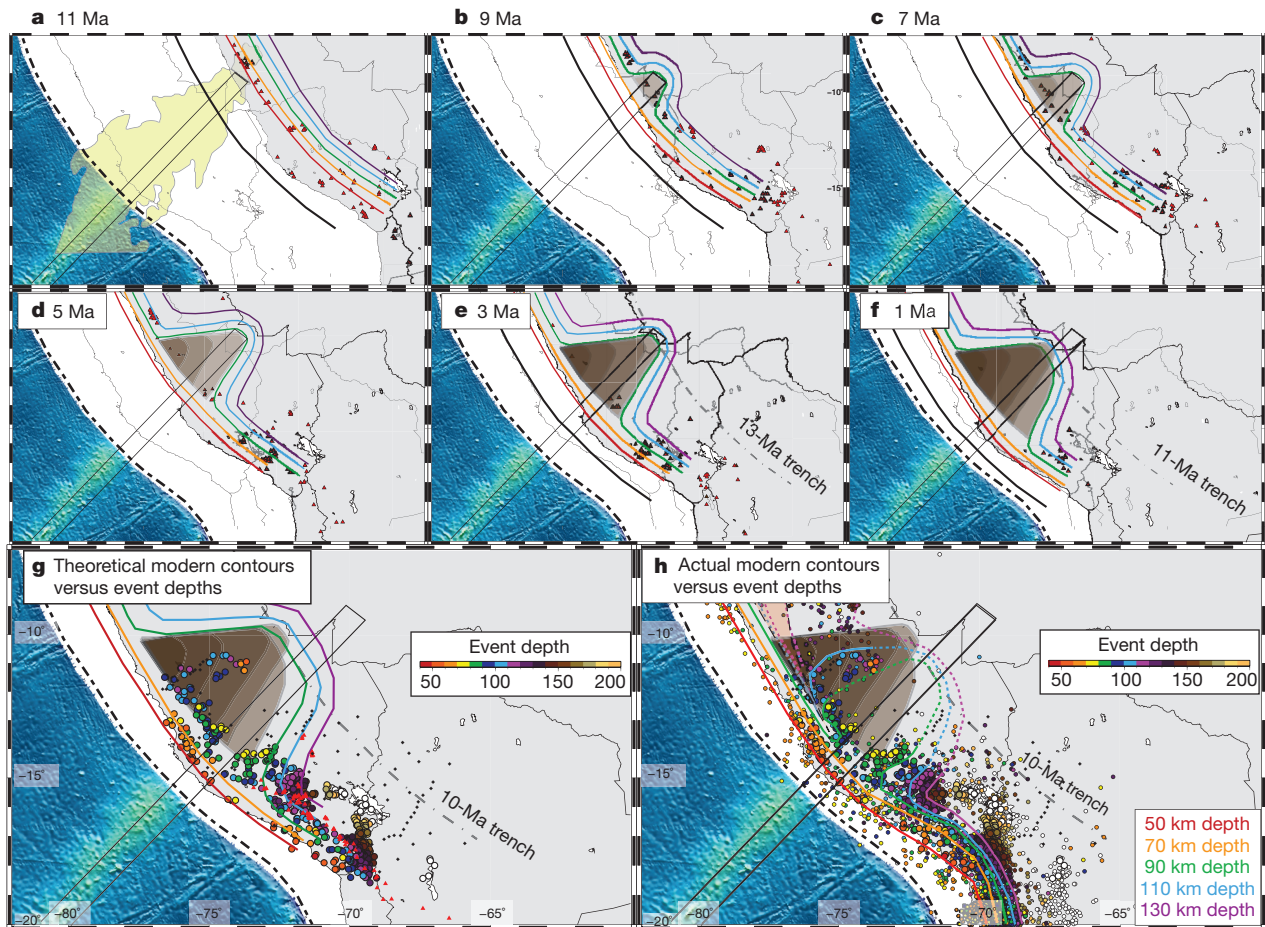


Figure 3 | Proposed evolution of the Peruvian flat slab. **a–f**, Proposed contours of the subducted slab, assuming that the ridge remains buoyant for 10 Ma after entering the trench. The approximate location of the subducted ridge is denoted by the black rectangular outline. Brown areas show areas of the continent underlain by flat slab at each time step. Triangles indicate volcanoes active during the 2 Myr following the time of the frame shown²². The location of the South American continent relative to the Nazca Ridge follows ref. 8. In **a**, we show the location of the projection of the mirror image of the Nazca Ridge (in yellow) that formed synchronously with the Nazca Ridge on the Pacific Plate when these plates were first created at the spreading centre following ref. 8. In **e**, red triangles show volcanism from 3 Ma to 2 Ma, and brown triangles show volcanism from 2 Ma to 1 Ma. In **f**, volcanism is shown

for 1 Ma to 0 Ma (not including Holocene volcanism). **g**, Modern seismicity from this study (large circles) with depths >50 km, and contours as they would be if the removal of the ridge did not affect the longevity of the flat slab. **h**, Modern seismicity from this study and local seismicity at depth >50 km, as reported in the ISC catalogue for years 2004–2014, shown as smaller circles¹⁷. We plot our observed slab contours on the basis of our earthquake locations and the location of high-velocity anomalies in our tomographic results. Dashed lines indicate contours that are less certain, either because of a paucity of earthquakes or because they lie outside of our region of good tomographic resolution. The pink triangular shape shows the region with very limited seismicity that may indicate a slab window caused by tearing and the reinitiation of normal subduction.

some portion of the Nazca Ridge has resumed normal subduction. We propose that, over time, the kinetically slow conversion of basalt and gabbro to eclogite in the overthickened crust of the Nazca Ridge results in an increase in the density of the horizontal plate. Given the inboard extent of the modern flat slab, we propose that, approximately 10 Ma after entering the trench, the overthickened oceanic crust of the Nazca Ridge becomes sufficiently eclogitized that it is no longer neutrally buoyant and therefore resumes its vertical descent (Figs 2b, g and 3e–g).

Finally, modelling studies indicate that suction between the horizontal plate and overriding continental lithosphere hinders the removal of the flat slab⁵. In our study area, this is important because the portion of the continent under which the flat slab initially forms moves northwest relative to the ridge over time. To test whether the flat slab will perpetuate beneath these continental regions after the departure of the ridge, we apply the fourth principle to our model: continental regions previously underlain by the flat slab will continue to have flat slab beneath them for some time (brown regions in Fig. 3). This results in a broadening of the flat slab as new continental areas to the south become underlain by the ridge and associated flat

slab, while areas to the north that were previously underlain by the ridge maintain their flat-slab geometry. This is consistent with earlier studies that attribute the along-strike (trench parallel) extent of the Peruvian flat slab to the southward sweep of the Nazca Ridge over time^{8,9}.

The proposed temporal evolution of the Peruvian flat slab shown in Fig. 3 combines the influences of trench retreat/overriding plate motion, suction, and ridge buoyancy. It assumes that the combination of all three forces is necessary for the formation of the flat slab, but that the first two are sufficient to perpetuate the flat slab after the departure of the ridge. A comparison between our conceptual model's slab geometry at present (Fig. 3g) with actual (observed) slab geometry (Fig. 3h) allows us to test these assumptions. The abrupt edge of the flat slab that we observe south of the ridge is very similar to that proposed by our conceptual model. We note that the dominant principle controlling the geometry of the flat slab here is the effect of ridge buoyancy, as there is no difference in trench rollback or continental lithospheric structure that might affect suction along strike in this region. Our observations therefore support the necessary contribution of the ridge to the formation of flat slabs, but are

also not inconsistent with additional contributions from suction and trench rollback.

Differences between the observed slab geometry and the geometry derived from our conceptual model are visible to the north of the ridge. In this area, the effect of the ridge is no longer present, and the geometry of the flat slab in our conceptual model is controlled by the effects of suction and trench rollback alone. Although both our conceptual model and our observations indicate a flat slab that broadens to the northwest of the ridge, the detailed morphologies are very different. In addition to an overall deepening of the flat slab north of the ridge (Fig. 3g, h and Extended Data Fig. 1), we observe a clear trench-parallel break in the subducted plate and a resumption of normal subduction trenchward of this tear (Figs 2a–c and 3h). This strongly suggests that, despite the presence of suction and trench rollback, the flat slab is no longer stable once the buoyant Nazca Ridge has been removed. Furthermore, once a break is present, the newly subducted plate assumes a normal steep dip angle, rather than a flat-slab geometry. In this study we are not able to resolve the northern extent of the Peruvian flat slab, nor can we establish the along-strike extent of the tear. However, International Seismological Centre Catalog locations north of our study area show a gap in seismicity that may be consistent with the absence of a flat slab because of a progressively tearing plate (Fig. 3h)¹⁷. The northward extent of the flat slab east of the tear may be due in part to the subduction of the Inca Plateau⁸, although this is beyond the scope of our study.

Our model is applicable to all flat-slab geometries in cases for which a distinct change of dip angle is observed. This change in dip occurs at the depth at which the slab becomes neutrally buoyant. Our results may not be applicable to slabs when the dip angle is constant but very shallow⁷ (for example, in Alaska and in the Cascadia region of the USA). Slabs that dip at a shallow angle sink at a constant rate, which is inconsistent with a period of neutral buoyancy. Such slabs may have effects that are similar to those produced by flat slabs, although they do not result in a complete cessation of arc volcanism (as occurred during the Laramide orogeny and is observed in Peru today), only its inboard deflection.

Our results may provide insights into the final stages of flat-slab subduction. Previous studies used volcanic patterns to reconstruct the formation and foundering of the Farallon flat slab in the western USA^{2,18,19}. The diversity of models for the progression of this foundering is indicative of the insufficiency of the constraints provided by volcanic trends alone. Our results suggest that once the flat slab extends some distance away from the buoyant feature, it will begin to sink and/or tear. Tearing of the Farallon plate caused by an excessively wide flat slab may be consistent with tomographic images of broken fragments of the Farallon plate²⁰.

We conclude that flat slabs form through a combination of trench retreat, suction, and the inability of overthickened oceanic crust to sink below some depth (about 90 km) until sufficiently eclogitized to become negatively buoyant once again. Flat slabs that extend laterally beyond some critical distance from the buoyant overthickened crust will begin to founder, even in the presence of other factors such as suction and trench retreat. The Peruvian flat slab provides insights into the temporal evolution of flat slabs from initial shallowing to collapse, yielding new constraints for the reconstruction of flat-slab genesis and the nature of the flat-slab foundering.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 October 2014; accepted 28 May 2015.

1. Bedle, H. & van der Lee, S. Fossil flat-slab subduction beneath the Illinois basin, USA. *Tectonophysics* **424**, 53–68 (2006).
2. Humphreys, E. *et al.* How Laramide-age hydration of North American lithosphere by the Farallon slab controlled subsequent activity in the western United States. *Int. Geol. Rev.* **45**, 575–595 (2003).
3. Vogt, P. R. Subduction and aseismic ridges. *Nature* **241**, 189–191 (1973).
4. van Hunen, J., Van Den Berg, A. P. & Vlaar, N. J. On the role of subducting oceanic plateaus in the development of shallow flat subduction. *Tectonophysics* **352**, 317–333 (2002).
5. Manea, V. C., Pérez-Gussinyé, M. & Manea, M. Chilean flat slab subduction controlled by overriding plate thickness and trench rollback. *Geology* **40**, 35–38 (2012).
6. Gerya, T. V., Fossati, D., Cantieni, C. & Seward, D. Dynamic effects of aseismic ridge subduction: numerical modelling. *Eur. J. Mineral.* **21**, 649–661 (2009).
7. Skinner, S. M. & Clayton, R. W. The lack of correlation between flat slabs and bathymetric impactors in South America. *Earth Planet. Sci. Lett.* **371–372**, 1–5 (2013).
8. Rosenbaum, G. *et al.* Subduction of the Nazca Ridge and the Inca Plateau: insights into the formation of ore deposits in Peru. *Earth Planet. Sci. Lett.* **239**, 18–32 (2005).
9. Eakin, C. M. *et al.* Response of the mantle to flat slab evolution: insights from local S splitting beneath Peru. *J. Geophys. Res.* **41**, 3438–3446 (2014).
10. Ward, K. M. *et al.* Ambient noise tomography across the Central Andes. *Geophys. J. Int.* **194**, 1559–1573 (2013).
11. Phillips, K. & Clayton, R. W. Structure of the subduction transition region from seismic array data in southern Peru. *Geophys. J. Int.* **196**, 1889–1905 (2014).
12. Cahill, T. & Isacks, B. L. Seismicity and shape of the subducted Nazca plate. *J. Geophys. Res.* **97**, 17503–17529 (1992).
13. Hayes, G. P., Wald, D. J., & Johnson, R. L. Slab 1.0: a three-dimensional model of global subduction zone geometries. *J. Geophys. Res.* **117**, B01302 (2012).
14. Hasegawa, A., & Sacks, I. S. Subduction of the Nazca plate beneath Peru as determined from seismic observations. *J. Geophys. Res.* **6**, 4971–4980 (1981).
15. Uyeda, S., Watanabe, T., Ozasayama, Y. & Ibaragi, K. Report of heat flow measurements in Peru and Ecuador. *Bull. Earthquake Res. Inst.* **55**, 55–74 (1980).
16. Arrial, P. A. & Billen, M. I. Influence of geometry and eclogitization on oceanic plateau subduction. *Earth Planet. Sci. Lett.* **363**, 34–43 (2013).
17. International Seismological Centre. *On-line Bulletin* (ISC, <http://www.isc.ac.uk> (2012)).
18. Liu, L. *et al.* The role of oceanic plateau subduction in the Laramide orogeny. *Nature Geosci.* **3**, 353–357 (2010).
19. Jones, C. H., Farmer, G. L., Sageman, B. & Zhong, S. Hydrodynamic mechanism for the Laramide orogeny. *Geosphere* **7**, 183–201 (2011).
20. Sigloch, K., McQuarrie, N. & Nolet, G. Two-stage subduction history under North America inferred from multiple-frequency tomography. *Nature Geosci.* **1**, 458–462 (2008).
21. Gripp, A. E. & Gordon, R. G. Young tracks of hotspots and current plate velocities. *Geophys. J. Int.* **150**, 321–361 (2002).
22. Instituto Geológico Minero y Metalúrgico. On-line catalog. (INGEMMET, <http://www.ingemmet.gob.pe> (2014)).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Clayton and P. Davies for providing the records from eight PERUSE stations. The PULSE experiment was supported by NSF grants EAR-0944184 (to L.S.W.), EAR-0943991 (to S.L.B.) and EAR-0943962 (to M.D.L.). The CAUGHT project was supported by NSF grants EAR-0908777 (to L.S.W.) and EAR-0907880 (to S.L.B.).

Author Contributions S.K.A. generated the tomographic model. L.S.W. developed the model of temporal evolution. A.K. provided earthquake locations. S.K.A., L.S.W. and A.K. developed the ideas and wrote the paper. S.L.B., M.D.L., G.Z., H.T. and C.C. contributed to data collection and paper editing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.K.A. (sknezevi@live.unc.edu).

METHODS

Earthquake locations. We use ANTELOPE software to auto-detect earthquakes, using a short-term average (STA) versus long-term average (LTA) trigger mechanism. The lengths of the STA and LTA moving time windows were chosen to be 1 s and 10 s respectively. After manually inspecting the waveforms, we selected 977 earthquakes out of 3,000 auto-detected events. We picked primary (P) and secondary (S, shear) wave arrival times for 673 slab events using SEISAN²³. The selected events have the following characteristics: all events are in the depth range 50–310 km; travel-time misfit is less than 1 s; data are well recorded at a minimum of ten stations with azimuthal gap $\leq 270^\circ$ (see Supplementary Information).

For relative locations we use the program HYPODD²⁴ (Extended Data Fig. 1). We calculate differential times between P and S phases recorded at a common station for each event pair separated by ≤ 40 km. This interevent distance was interactively chosen after optimizing the linkage between the events in the first-step processing of phase data in HYPODD. Each event is strongly linked to a maximum of ten neighbouring events, having at least eight travel-time observations. We used the P-wave velocity model of ref. 25 for our starting model, and set the crustal thickness to 65 km. We used a V_p/V_s ratio of 1.75 to calculate S-wave velocities (where V_p is the P-wave velocity and V_s is the S-wave velocity).

Three-dimensional shear-wave imaging. The three-dimensional imaging of shear-wave velocity structure using earthquake-generated Rayleigh waves proceeds in two steps: first, we invert for Rayleigh wave phase velocities; subsequently, we invert the obtained phase velocities for shear-wave velocities. We use the two-plane wave method²⁶ to invert for Rayleigh wave phase velocities. Observations are modelled as a sum of two interfering plane waves, each described by its amplitude, phase and backazimuth. Predicted phase and amplitude values are calculated using finite frequency sensitivity kernels²⁷ that incorporate the (Born) single scattering approximation²⁸. Amplitudes are corrected for geometrical spreading and attenuation. We examined 12 periods in the band between 0.007 Hz and 0.03 Hz, sensitive to V_s structure from the lower crust (~ 40 -km depth), to the upper mantle (~ 200 -km depth).

Data were collected from several seismic networks: PULSE⁹, CAUGHT¹⁰, PERUSE¹¹, and the global network permanent station in Lima, Peru. We picked fundamental mode Rayleigh waves for 65 well recorded teleseismic events (Extended Data Fig. 2a) with magnitudes ≥ 5.5 .

We defined the study area with corners at 10° S, 69° W; 18° S, 79° W; 10° S, 69° W; and 10° S, 69° W (Extended Data Fig. 2b). The starting velocity model (Extended Data Fig. 2c) accounts for different crustal thicknesses across the study area²⁹. We combine the IASPEI91 velocity model for the mantle³⁰ and the model of ref. 31 for the crust and use a forward algorithm³² to predict phase velocities across the study region.

The inversion is regularized with model covariances set to 0.15 km s^{-1} . The choice of regularization parameter is based on the stability of both Rayleigh-wave and shear-wave inversions. Longer periods are generally less well resolved than shorter periods because of their broader sensitivity kernels. The best resolved areas are beneath the Western Cordillera, Altiplano, Eastern Cordillera, and coastal forearc and, to a lesser extent, the Sub-Andean zone (Extended Data Fig. 3a). The resolution within the foreland basin is mostly confined along the stations deployed in foreland basin in eastern Peru.

In the second step, we invert obtained phase velocities (Extended Data Fig. 4) for one-dimensional shear-wave velocities^{32,33}. We use the same starting model as in the previous step (Extended Data Fig. 2c). Sensitivity kernels for longer periods are substantially broader than the sensitivity kernels for shorter periods, and sample greater depths. The peak sensitivities for the periods used in this study range from depths of ~ 40 km (for 33 s) up to ~ 200 km (for 143 s). Thus, the vertical resolution is greatest between ~ 40 km and ~ 200 km, and decreases gradually with depth (below ~ 300 km, resolution drops below 0.1; Extended Data Fig. 3b). The model covariance obtained for phase velocities from the two-plane wave method was used as data covariance to regularize the shear wave velocity inversion. The average root mean squared (r.m.s.) misfit between predicted and observed phase velocities over all periods indicates an average error of $\sim 0.02 \text{ km s}^{-1}$ (Extended Data Fig. 3c). Results of our shear-wave inversions are presented in Fig. 2 and Extended Data Fig. 5.

Lateral and vertical resolution. The main new features observed in this study from the surface-wave tomography include the far inboard extent of flat slab along the subducting Nazca ridge, and the slab tear north of the ridge. We performed a range of tests to investigate lateral and vertical resolution to ensure the robustness of these features (see also Extended Data Figs 3, 6–10).

Lateral resolution. We plot the resolution matrix rows of isolated model parameters for several periods, with an emphasis on the spatial resolution at three locations along the northern profile north of the subducting Nazca Ridge: one where we observe re-steepening of the slab, one at the slab tear, and one along the flat-slab remnant. We also investigate points at two locations along the subducting

Nazca Ridge: one where we observe the far inboard extent of the flat slab ('long flat slab'), and one where previous studies¹² suggest the end of flat slab should be ('short flat slab'). The examination of our resolution matrix for these five selected nodes is intended primarily to demonstrate that we have sufficient spatial resolution to resolve the slab tear north of the ridge and the inboard extent of flat slab along the ridge. We focus on intermediate periods because they have peak sensitivity at the most relevant depths (Extended Data Fig. 2c). The tests show that these model parameters are able to resolve spatial-scale features smaller than those discussed here. The only node for which we observe a particularly broad sensitivity cone is the one at the far inboard extent of the flat slab. This finding suggests that, while the inboard extent of the flat slab may not be as well resolved as in other locations, a shorter flat slab would have been imaged accurately if it did exist. Our inboard extent is therefore a conservative estimate.

To demonstrate the sensitivity of our results to grid node spacing, we plot phase-velocity maps for intermediate periods using 0.25° and 0.5° grid node spacing. The phase velocity maps with 0.5° grid node spacing show major features that are smoother than, but consistent with, major features that are observed on maps with 0.25° spacing. Further, the dispersion curves for the five selected nodes reflect consistency regardless of the grid node spacing. Along the northern profile in both cases we observe faster anomalies at 66 s and 77 s, where we observe the re-steepened slab, slow anomalies at all intermediate periods where we observe the slab tear, and fast anomalies where we observe the flat-slab remnant. Along the flat-slab profile we note low phase velocities at the location where previous studies suggest a resumption of the steep slab, and high phase velocities at the location which we propose to be the end of flat slab.

We perform a series of checkerboard tests using the surface wave resolution matrices to test the size of the anomalies that can be recovered with the varying periods used here (Extended Data Fig. 6). These tests show whether we have sufficient spatial resolution to recover the size of the anomaly analogous to the observed tear and whether we have sufficient resolution to resolve the inboard extent of the flat slab. For this reason we plot the five selected nodes. In addition, these tests yield a better understanding of the spatial resolution of phase velocity maps across the study area and easily reveal areas that suffer from smearing (because of preferential ray path direction and/or lack of data). Short and intermediate periods, with peak sensitivities between 50-km and 150-km depth, are able to recover smaller anomalies, equal to and smaller than the lateral extent of the observed slab tear. The tests show that we do have sufficient spatial resolution to resolve the slab tear, flat slab remnant to the east, and re-steepened slab to the west. Longer periods, which mostly sample subslab material, can recover slightly larger features. However, both shorter and longer periods are able to resolve the size of the anomaly analogous to subducting slab at the end of the flat slab. These checkerboard tests show that we are able to resolve anomalies where previous studies suggested the end of flat slab, while the node representing the far inboard extent of flat slab may be streaked because of a lack of crossing rays. Thus, on the basis of these tests, we can conclude with confidence that the inboard extent of flat slab along the subducting Nazca Ridge is not where previously assumed, but further inboard. Resolution at the location representing the far inboard extent of flat slab is weak and suffers from smearing. However, our conclusion on the far inboard extent of flat slab is supported by constraints from other studies^{34,35}.

Vertical resolution. We test our vertical resolution along all profiles shown in Fig. 2. Extended Data Figures 7–10 show recovery tests for: the southern profile, where we observe steeply dipping slab (Extended Data Fig. 7); the flat-slab segment along the Nazca Ridge (Extended Data Fig. 8); just north of the Nazca Ridge, where we observe deepening of earthquakes and the start of slab tear (Extended Data Fig. 9); and the northern profile, where we observe the slab tear, re-steepening of the slab and flat-slab remnant (Extended Data Fig. 10).

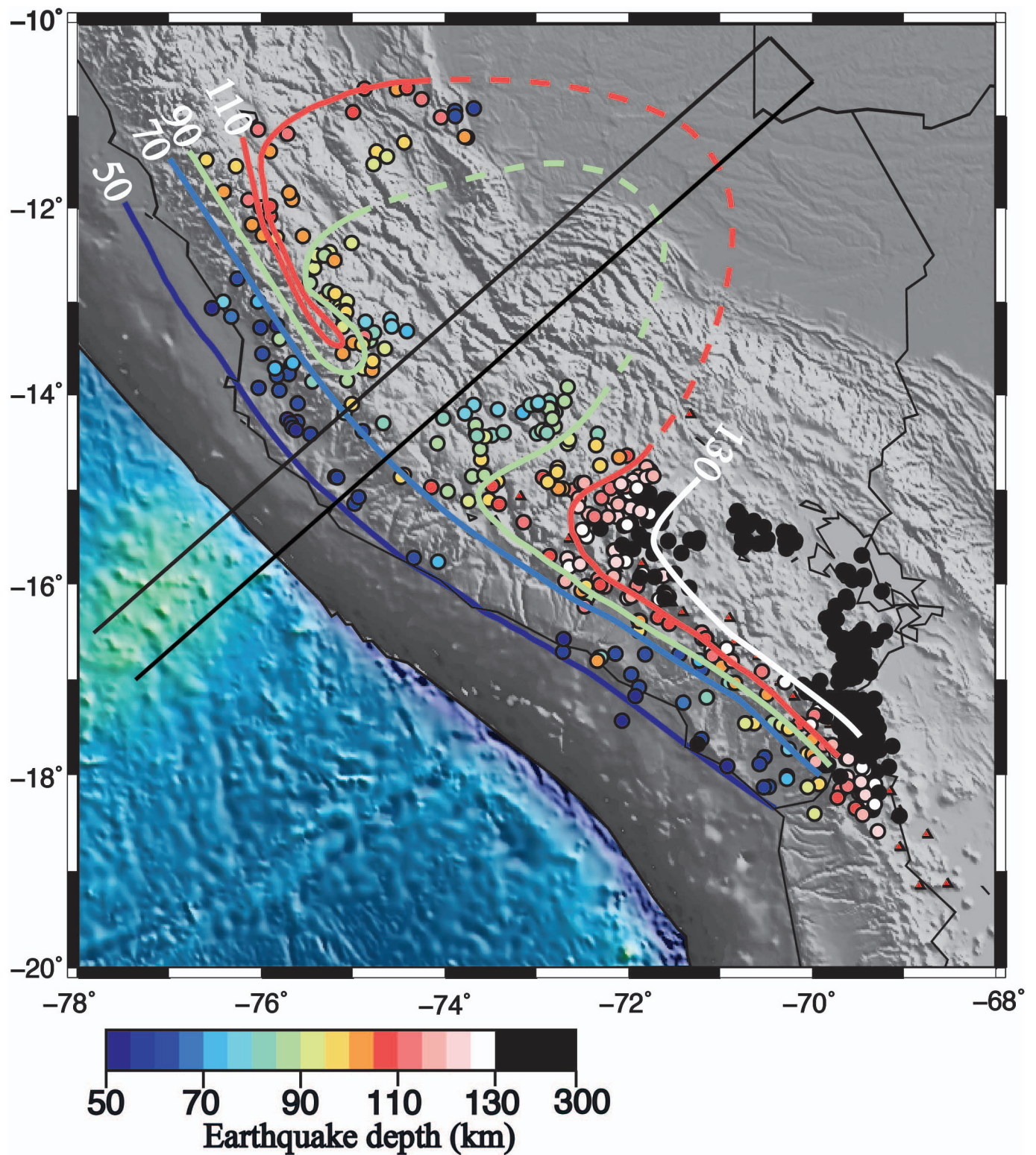
Extended Data Fig. 7 demonstrates our ability to recover a dipping slab south of the ridge. We model a shear wave velocity structure with a 70-km-thick steeply dipping slab associated with a velocity of 4.6 km s^{-1} (Extended Data Fig. 7d). This model is based on our interpretations of the observed structures that we show in Fig. 2d. We predict dispersion curves for this model using the code of ref. 32, add noise to predicted phase velocities, and invert them using the same starting model (Extended Data Fig. 7b) and regularization parameters as for the model shown in Extended Data Fig. 7c. The Gaussian noise was generated from misfits obtained in our final model using the central limit theorem method, and randomly assigned to predicted phase velocities. We were able to recover the steeply dipping structure, but its thickness appears greater owing to vertical smearing. We were not able to recover the full amplitude of the anomaly, but a somewhat lower amplitude (4.45 – 4.55 km s^{-1}). Our model calculated using observed data (Extended Data Fig. 7c) indicates shear-wave velocities above 4.55 km s^{-1} . This recovery test suggests that, in order to fully recover the amplitude of observed high shear-wave velocities, either the slab in Extended Data Fig. 7d needs to be associated with velocities greater than 4.6 km s^{-1} , or the thickness of the slab should be greater, or both.

Extended Data Figure 8 demonstrates our ability to differentiate between a flat slab with our (greater) inboard extent along the Nazca ridge track ('long flat slab') and a flat slab with shorter extent, suggested previously¹² (Extended Data Fig. 8g, 'short flat slab'). The plots in Extended Data Fig. 8h and i show recovered models. The tests show that we are able to recover the flat-slab-related high shear-wave velocities. However, we observe vertical smearing, and notice that, in the recovered model, slab-related high velocities appear at shallower depths, resulting in high velocities in the lower crust and more shallow flat slab. This is also noticeable in our model shown in Extended Data Fig. 8e. Owing to vertical smearing and the gradational nature of the slab–mantle boundary in oceanic plates, the bottom of the slab is poorly resolved. Different layer discretization owing to different crustal thicknesses causes the artificial undulating nature of the slab's positive anomaly, also present in our model (Extended Data Fig. 8e). The plots in Extended Data Fig. 8h, i demonstrate sufficient vertical resolution to recover the end of the flat slab. We also plot dispersion curves at two points, representing the shorter end of flat slab previously suggested¹² (Extended Data Fig. 8b) and the far inboard end (Extended Data Fig. 8c). Dispersion curves predicted for shorter and longer flat slab are substantially different and the observed phase velocities match better with the longer flat slab.

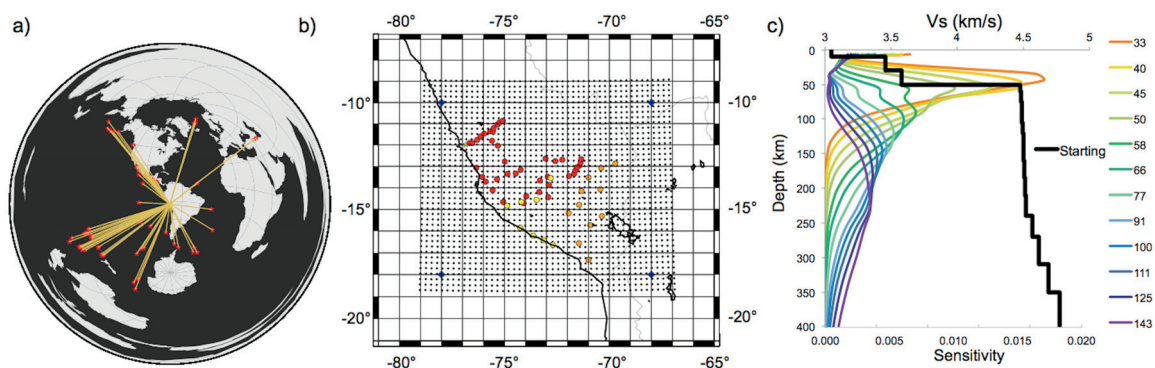
Extended Data Figure 9 demonstrates our ability to recover a torn slab to the north of the Nazca ridge. Extended Data Figure 10 demonstrates our ability to distinguish between torn slab and a continuous slab along the northernmost profile. Extended Data Figure 10i, j shows that lateral heterogeneities and dipping structures are well recovered (except at shallower depths, where we lose resolution; Extended Data Fig. 3b). Again, we are able to recover the flat slab, but with evident vertical smearing. The observed dispersion curves at locations at which we observe the re-steepened slab (point 1), torn slab (point 2) and flat-slab remnant (point 3) are very different. Shorter periods of the torn slab model at point 1 are characterized with low phase velocities, while intermediate periods have much higher phase velocities. In contrast, the continuous slab model is associated with high phase velocities at both short and intermediate periods. At point 2 both short and intermediate periods show low phase velocities for the torn slab model, but high phase velocities for the continuous slab model. At point 3 both short and intermediate periods are associated with high phase velocities for both torn and

continuous flat slab. Generally, we are able to reproduce the observed dispersion curves with our model of torn slab (Fig. 2d), except for the low phase velocities at shorter periods at point 1. This is because we did not introduce low shear velocities in the lower crust in our starting model. Dispersion curves for the continuous flat slab differ from those observed at points 1 and 2, especially at intermediate periods that sample upper mantle material.

23. Lienert, R. B., Berg, E. & Frazer, L. N. HYPOCENTER: an earthquake location method using centered, scaled, and adaptively damped least squares. *Bull. Seismol. Soc. Am.* **76**, 771–783 (1986).
24. Waldhauser, F. & Ellsworth, W. L. A double-difference earthquake location algorithm: method and application to the Northern Hayward Fault, California. *Bull. Seismol. Soc. Am.* **90**, 1353–1368 (2000).
25. Dorbath, C. Velocity structure of the Andes of central Peru from locally recorded earthquakes. *J. Geophys. Res.* **23**, 205–208 (1996).
26. Forsyth, D. W. & Li, A. in *Seismic Earth: Array of Broadband Seismograms* (eds Levander, A. & Nolet, G.) 81–98 (Geophysical Monograph Series 157, 2005).
27. Yang, Y. & Forsyth, D. W. Regional tomographic inversion of the amplitude and phase of Rayleigh waves with 2-D sensitivity kernels. *Geophys. J. Int.* **166**, 1148–1160 (2006).
28. Zhou, Y., Dahlen, F. A. & Nolet, G. Three-dimensional sensitivity kernels for surface wave observables. *Geophys. J. Int.* **158**, 142–168 (2004).
29. Tassara, A., Götze, H.-J., Schmidt, S. & Hackney, R. Three-dimensional density model of the Nazca plate and the Andean continental margin. *J. Geophys. Res.* **111**, B09404 (2006).
30. Kennett, B. L. N. *IASPEI 1991 Seismological Tables* (Bibliotech, 1991).
31. James, D. E. Andean crustal and upper mantle structure. *J. Geophys. Res.* **76**, 3246–3271 (1971).
32. Saito, M. in *Seismological Algorithms: Computational Methods and Computer Programs* (ed. Doornbos, D. J.) 293–319 (Elsevier, 1988).
33. Weeraratne, D. S., Forsyth, D. W., Fischer, K. M., & Nyblade, A. A. Evidence for an upper mantle plume beneath the Tanzanian craton from Rayleigh wave tomography. *J. Geophys. Res.* **108**, 2427 (2003).
34. Scire, A. *et al.* Imaging the transition from flat to normal subduction: variations in the structure of the Nazca slab and upper mantle under southern Peru and northwestern Bolivia. *Geophys. J. Int.* (submitted).
35. Snoke, J. A., Sacks, I. S. & Okada, H. Determination of the subducting lithosphere boundary by use of converted phases. *Bull. Seismol. Soc. Am.* **67**, 1051–1060 (1977).

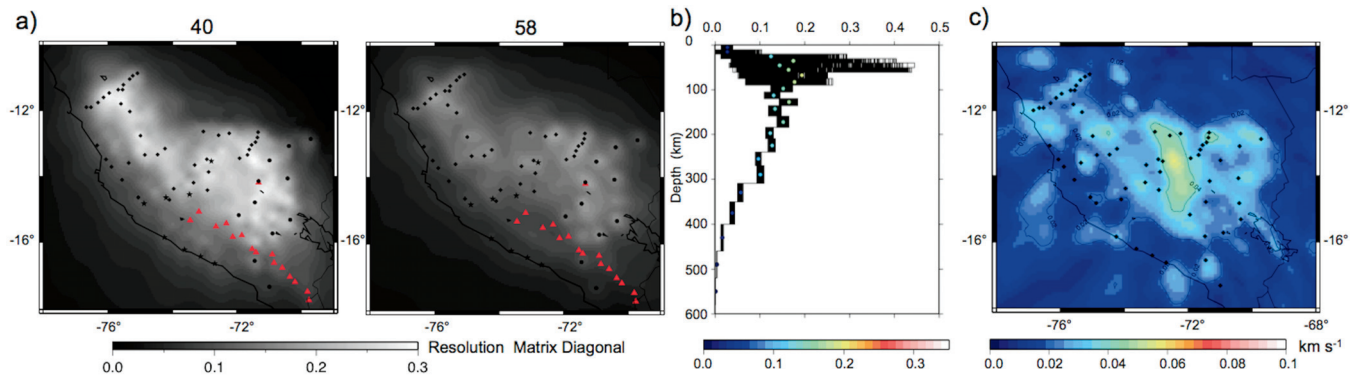


Extended Data Figure 1 | Relocated earthquakes used in this study. Colours show the depth of events (in km) and lines indicate the slab contours in 20-km depth increments. Events below 130 km are shown in black.



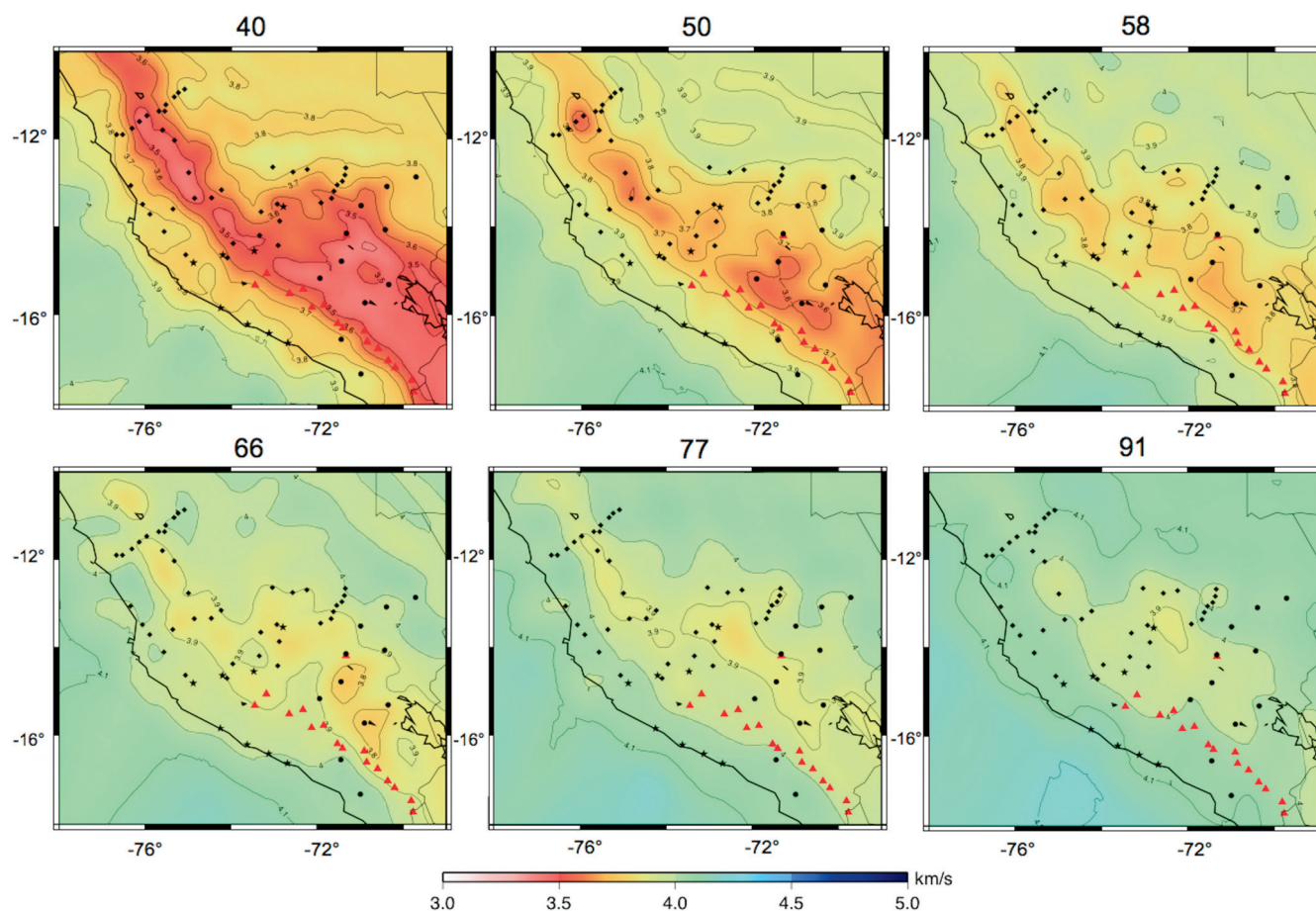
Extended Data Figure 2 | Events, grid and starting model used for the Rayleigh wave phase velocity inversions. **a**, Teleseismic events used in the study. **b**, Black diamonds represent grid nodes; blue diamonds represent corners used in the two-plane wave methodology; red circles are PULSE

stations; orange circles are CAUGHT stations; yellow circles are PERUSE stations; and yellow star is the permanent NNA station. **c**, Sensitivity kernels for periods used in our study with the one-dimensional starting shear wave velocity model.

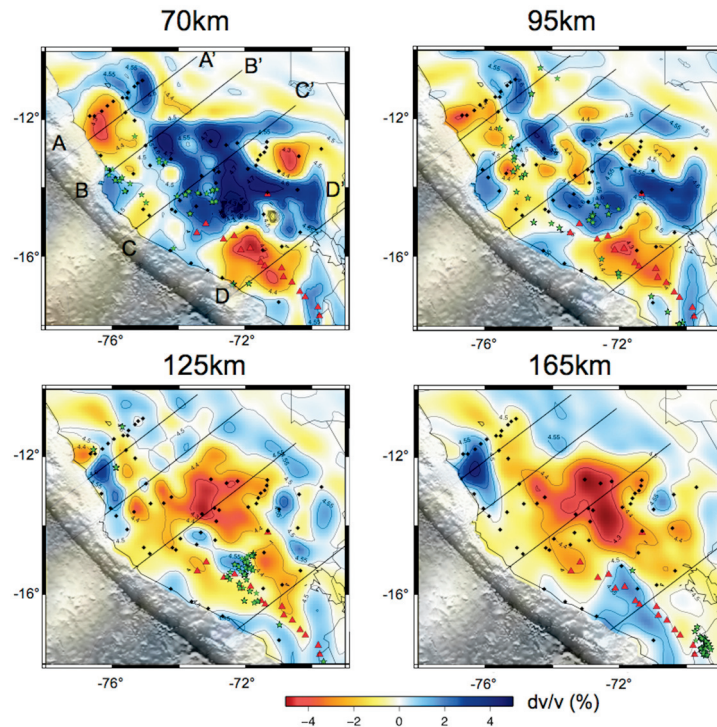


Extended Data Figure 3 | Lateral and vertical resolution. **a**, Resolution for the 40-s and 58-s periods. The resolution matrix diagonal for Rayleigh wave phase velocities is indicated in grey scale. Red triangles are Holocene volcanoes. Black rectangles, circles and stars are stations used here. **b**, Resolution matrix diagonal values for all one-dimensional shear wave velocity inversions.

Colours of the circles indicate average values of the resolution matrix diagonals for each layer. **c**, The r.m.s. average misfit over all periods at each point after our shear wave inversions. Colours represent the misfit in km s^{-1} . Black rectangles represent stations used in our study.

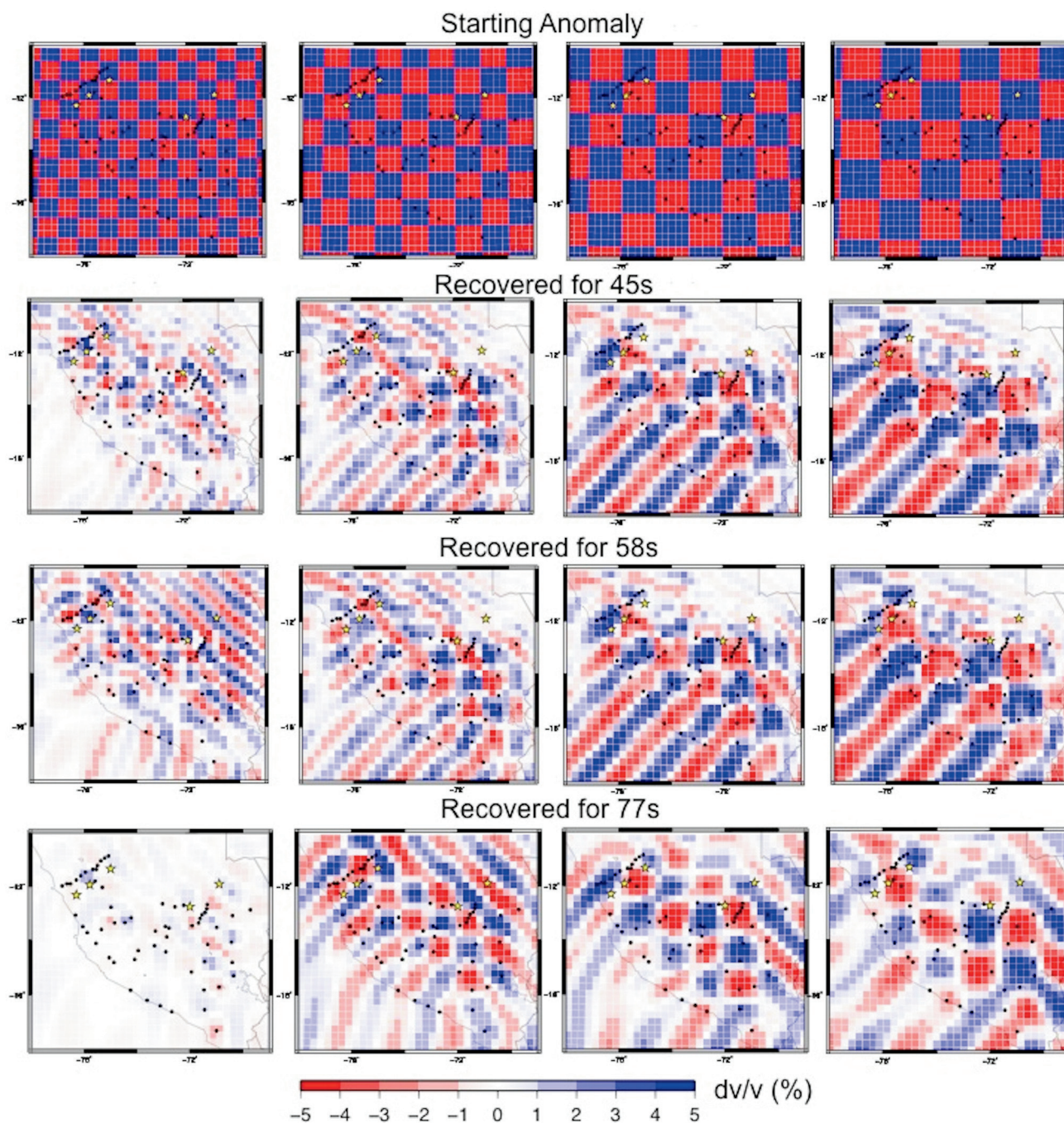


Extended Data Figure 4 | Calculated Rayleigh wave phase velocities for 40-s, 50-s, 58-s, 66-s, 77-s and 91-s periods. Colours and contours indicate absolute phase velocities. Red triangles represent Holocene volcanoes. Black rectangles, circles and stars represent stations used in our study.



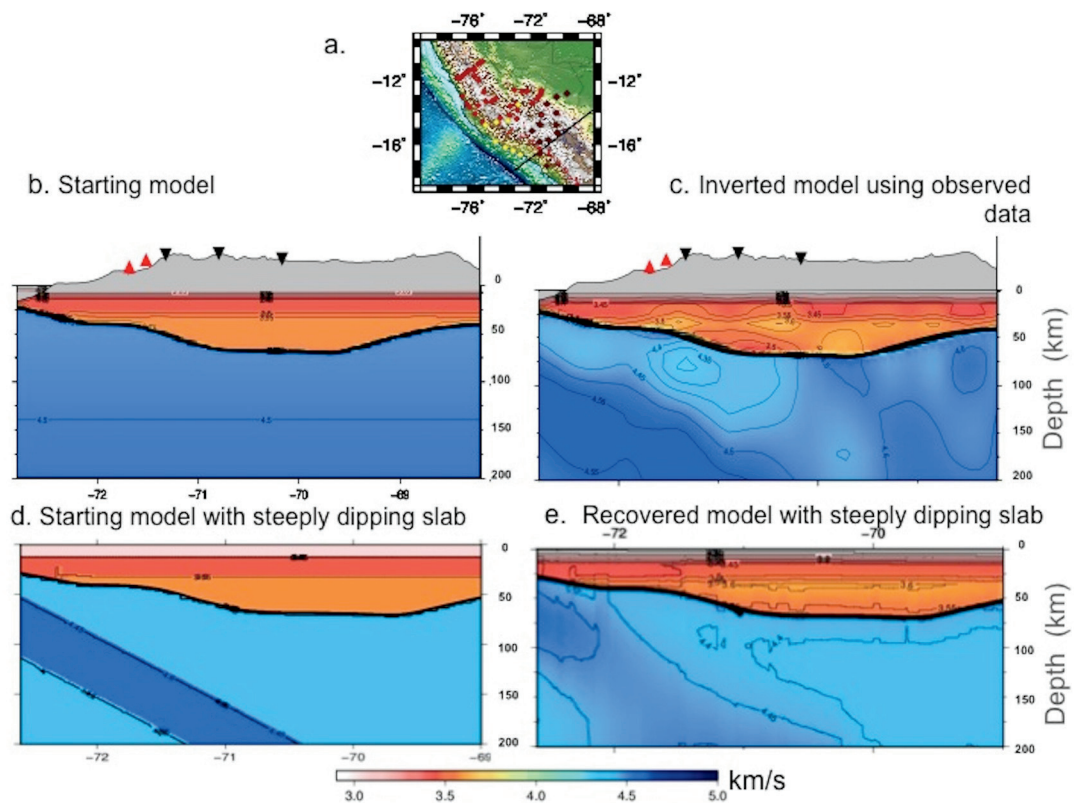
Extended Data Figure 5 | Shear wave velocity maps at depths of 70 km, 95 km, 125 km and 165 km. Colours represent velocity deviations with respect to the reference model (Extended Data Fig. 2c); contours show absolute

velocities. Green stars indicate earthquakes within a depth of 20 km relocated using HypoDD²⁴. Red triangles represent Holocene volcanoes. Black rectangles, circles and stars are stations used in our study.



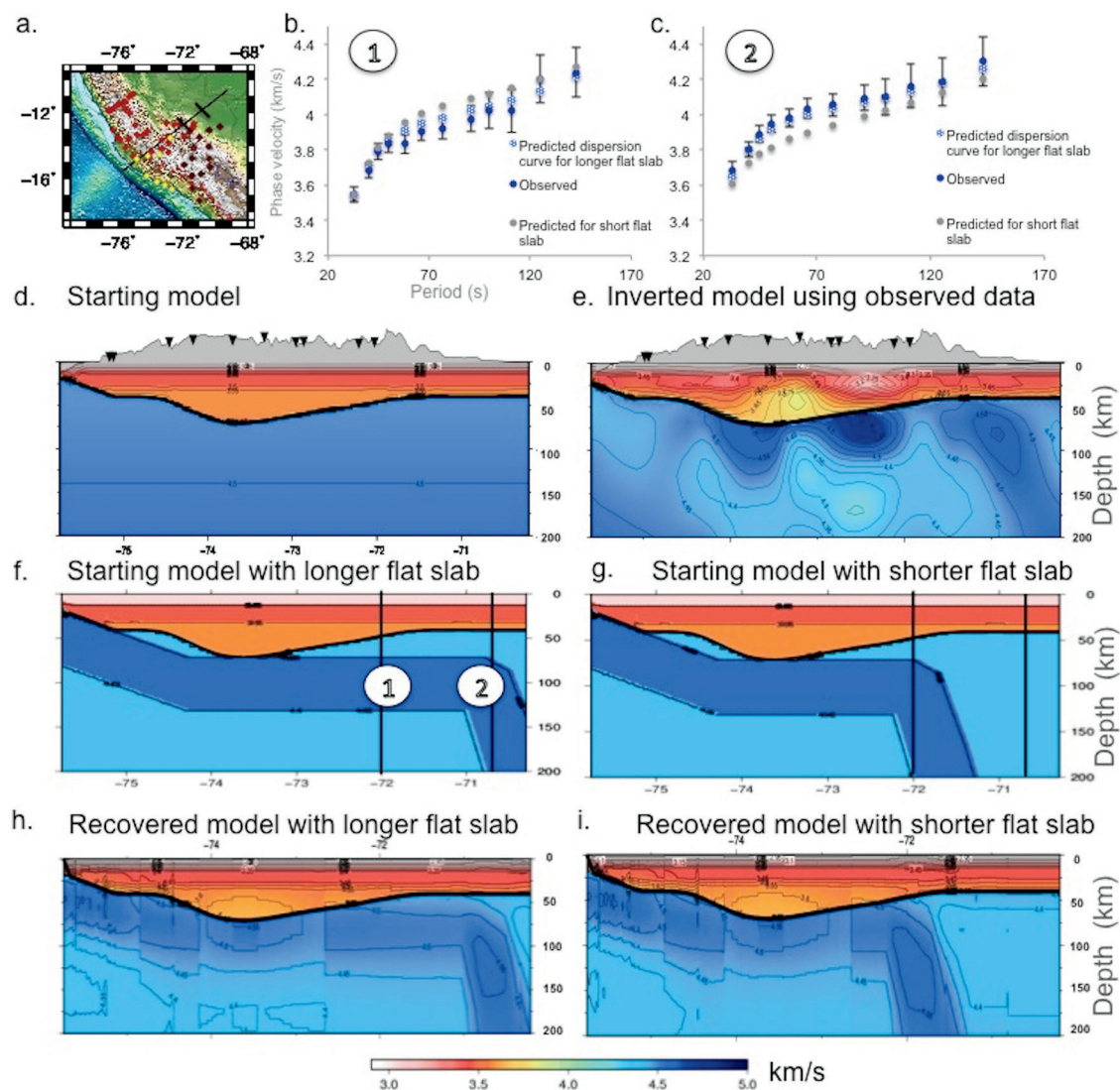
Extended Data Figure 6 | Checkerboard tests estimated from resolution matrix for 45 s, 58 s and 77 s. Colours represent the recovered anomaly. Yellow stars along the northernmost profile indicate locations where we observe re-steepening of the slab, slab tear and the flat-slab remnant; the two

stars along the subducting Nazca Ridge refer to locations of our (greater) inboard extent along the Nazca ridge track and a flat slab with shorter extent suggested previously¹².



Extended Data Figure 7 | Recovery tests for the dipping slab south of the ridge. **a**, Map showing the transect and stations locations (rectangles). **b**, Starting model used in the shear wave velocity inversion; contour lines and colours are absolute shear-wave velocities; black inverted triangles are

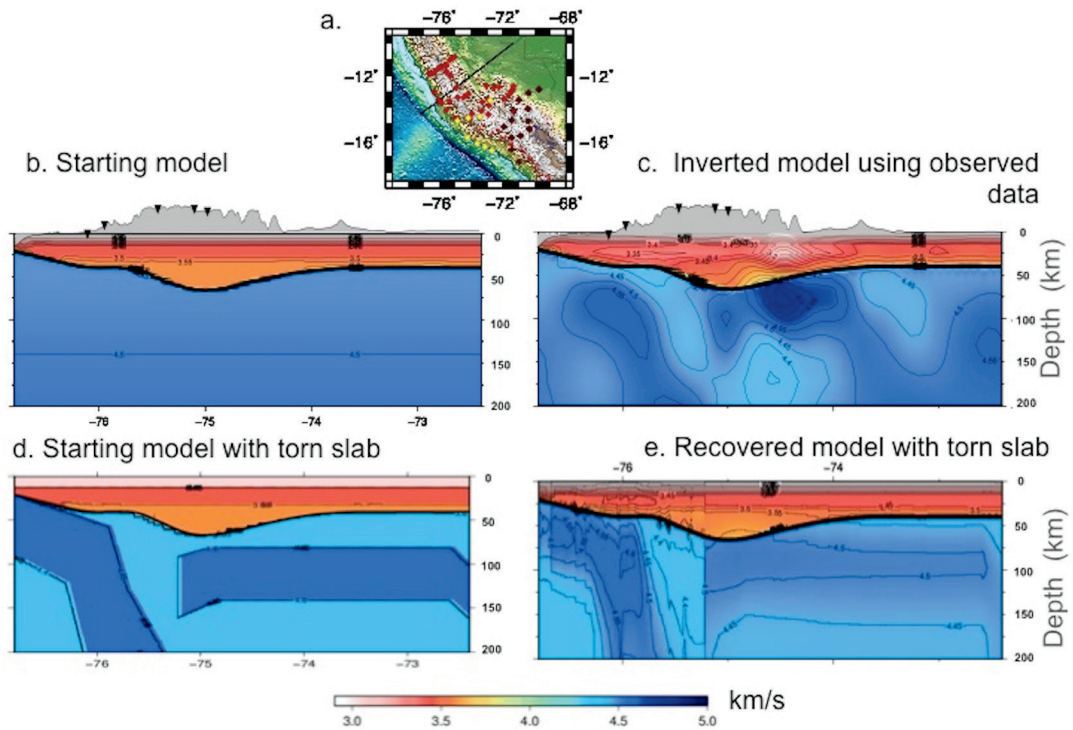
stations within 35 km of the transect; red triangles are Holocene volcanoes. **c**, Model calculated using observed data. **d**, Model based on our interpretations of the observed structures (see also Fig. 2d). **e**, Recovered model.



Extended Data Figure 8 | Recovery tests for the flat-slab segment.

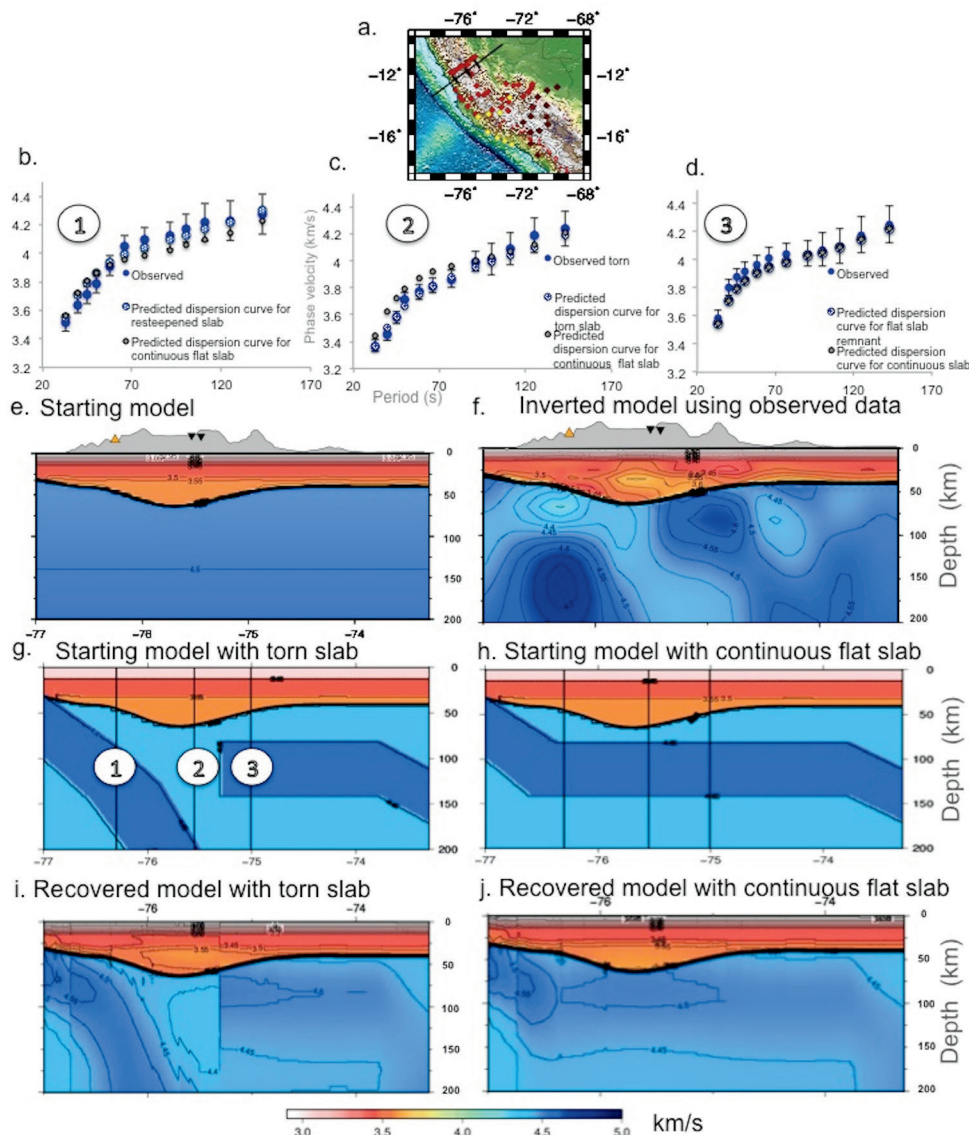
a. Transect. **b.** Dispersion curve at a location representing the shorter end of flat slab previously suggested¹². **c.** Dispersion curve at a location representing the greater inboard extent of flat slab (proposed here); error bars represent one

standard deviation of uncertainty. **d.** Starting model. **e.** Model calculated using observed data. **f.** Model with our (greater) inboard extent of flat slab. **g.** Model with shorter flat slab as suggested previously¹². **h.** Recovered model from **f.** **i.** Recovered model from **g.**



Extended Data Figure 9 | Recovery tests for the area just north of the Nazca Ridge, where we observe a deepening of earthquakes, and the southern end of the slab tear. **a**, Reference map. **b**, Starting model. **c**, Model calculated

using observed data. **d**, Model based on our interpretations of the torn slab (see Fig. 2d). **e**, Recovered model.



Extended Data Figure 10 | Recovery tests for the northern profile where we observe the slab tear, a re-steepening of the currently subducting slab west of the tear, and the flat slab remnant east of the tear. a, Transect. b, Dispersion curve at the location representing the re-steepened slab; error bars represent one standard deviation of uncertainty. c, Dispersion curve at location representing the slab tear. d, Dispersion curve at location representing

the flat slab remnant. e, Starting model. f, Model calculated using observed data. g, Model with slab tear that we propose in this study. h, Model with continuous flat slab suggested previously suggested in ref. 12 and other studies. i, Recovered model from g. j, Recovered model from h. See Methods for further explanation.

An early modern human from Romania with a recent Neanderthal ancestor

Qiaomei Fu^{1,2,3*}, Mateja Hajdinjak^{3*}, Oana Teodora Moldovan⁴, Silviu Constantin⁵, Swapan Mallick^{2,6,7}, Pontus Skoglund², Nick Patterson⁶, Nadin Rohland², Iosif Lazaridis², Birgit Nickel³, Bence Viola^{3,7,8}, Kay Prüfer³, Matthias Meyer³, Janet Kelso³, David Reich^{2,6,9} & Svante Pääbo³

Neanderthals are thought to have disappeared in Europe approximately 39,000–41,000 years ago but they have contributed 1–3% of the DNA of present-day people in Eurasia¹. Here we analyse DNA from a 37,000–42,000-year-old² modern human from Peștera cu Oase, Romania. Although the specimen contains small amounts of human DNA, we use an enrichment strategy to isolate sites that are informative about its relationship to Neanderthals and present-day humans. We find that on the order of 6–9% of the genome of the Oase individual is derived from Neanderthals, more than any other modern human sequenced to date. Three chromosomal segments of Neanderthal ancestry are over 50 centimorgans in size, indicating that this individual had a Neanderthal ancestor as recently as four to six generations back. However, the Oase individual does not share more alleles with later Europeans than with East Asians, suggesting that the Oase population did not contribute substantially to later humans in Europe.

Between 45,000 and 35,000 years ago, anatomically modern humans spread across Europe, while the Neanderthals, present since before 300,000 years ago, disappeared. How this process occurred has long been debated^{1,3–5}. Comparisons between the Neanderthal genome and the genomes of present-day humans have shown that Neanderthals contributed approximately 1–3% of the genomes of all people living today outside sub-Saharan Africa^{6,7} suggesting that human populations ancestral to all non-Africans mixed with Neanderthals. The size of segments of Neanderthal ancestry in present-day humans suggests that this occurred between 37,000 and 86,000 years ago⁸. However, where and how often this occurred is not understood. For example, Neanderthals share more alleles with East Asians and Native Americans than with Europeans, which may reflect additional interbreeding in the ancestors of eastern non-Africans^{9–12}. Surprisingly, analyses of present-day genomes have not yielded any evidence that Neanderthals mixed with modern humans in Europe, despite the fact that Neanderthals were numerous there and cultural interactions between the two groups have been proposed^{13,14}.

More direct insight into the interactions between modern and archaic humans can be obtained by studying genomes from modern humans who lived at a time when they could have met Neanderthals. Recent analyses of genomes from a ~43,000–47,000-year-old modern human from western Siberia¹⁵ and a ~36,000–39,000-year-old modern human from eastern Europe¹⁶ showed that Neanderthal gene flow into modern humans occurred before these individuals lived. The Siberian individual's genome contained some segments of Neanderthal ancestry as large as 6 million base pairs (bp), suggesting that some Neanderthal gene flow could have occurred a few thousand years before his death¹⁵.

We report genome-wide data from a modern human mandible, Oase 1, found in 2002 in the Peștera cu Oase, Romania. The age of this specimen has been estimated to be ~37,000–42,000 years by direct radiocarbon dating^{2,17,18}. Oase 1 is therefore one of the earliest modern humans in Europe. Its morphology is generally modern but some aspects are consistent with Neanderthal ancestry^{19–21}. Subsequent excavations uncovered a cranium from another, probably contemporaneous individual, Oase 2, which also carries morphological traits that could reflect admixture with Neanderthals^{17,19}.

We prepared two DNA extracts from 25 mg and 10 mg of bone powder removed from the inferior right ramus of Oase 1. We treated an aliquot of each of these extracts with *Escherichia coli* uracil-DNA glycosylase (UDG), an enzyme that removes uracils from the interior parts of DNA molecules, but leaves a proportion of uracils at the ends of the molecules unaffected. Uracil residues occur in DNA molecules as a result of deamination of cytosine residues, and are particularly prevalent at the ends of ancient DNA molecules^{9,22}. Among the DNA fragments sequenced from these two extracts, 0.18% and 0.06%, respectively, could be mapped to the human reference genome. We prepared three additional DNA libraries from the extract containing 0.18% human-like molecules, but omitted the UDG treatment to increase the number of molecules in which terminal C-to-T substitutions could be seen and used to identify putatively ancient fragments. Because the fraction of endogenous DNA is so small, we used hybridization to DNA probes to isolate human DNA fragments from the libraries²³. Applying this strategy to the mitochondrial genome allowed the mitochondrial (mt)DNA from the five libraries to be sequenced to an average coverage of 803-fold (Supplementary Note 1). At the 3' ends of the DNA fragments, cytosine residues appeared as thymine residues relative to the human mtDNA reference in 21% of fragments, reflecting appreciable levels of cytosine deamination. This suggests that at least some of the human mtDNA is of ancient origin. We determined mtDNA consensus sequences in two ways: using all mtDNA fragments, and using only deaminated fragments that carry C-to-T substitutions at either end relative to the consensus mtDNA sequence based on these fragments, an approach known to enrich for endogenous DNA^{9,24–26}. The mtDNA sequence based on all fragments clusters with present-day Europeans (Extended Data Fig. 1) (Supplementary Note 1). In contrast, the mtDNA sequence based on deaminated fragments is related to a large group of present-day Eurasian mtDNAs (haplogroup N) but diverges from these before they diverged from each other. This Oase 1 mtDNA carries a few private mutations on the basis of which its age can be estimated to be 36,330 years before present (14,520–56,450; 95% confidence interval). Using six positions at which the mtDNA sequence differs from at least 99% of 311 present-day humans, we estimate the contamination

¹Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China. ²Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. ⁴“Emil Racoviță” Institute of Speleology, Cluj Branch, 400006 Cluj, Romania. ⁵“Emil Racoviță” Institute of Speleology, Department of Geospeleology and Paleontology, 010986 Bucharest 12, Romania. ⁶Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁷Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. ⁸Department of Anthropology, University of Toronto, Toronto, Ontario, M5S 2S2, Canada. ⁹Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, 02115, USA.

*These authors contributed equally to this work.

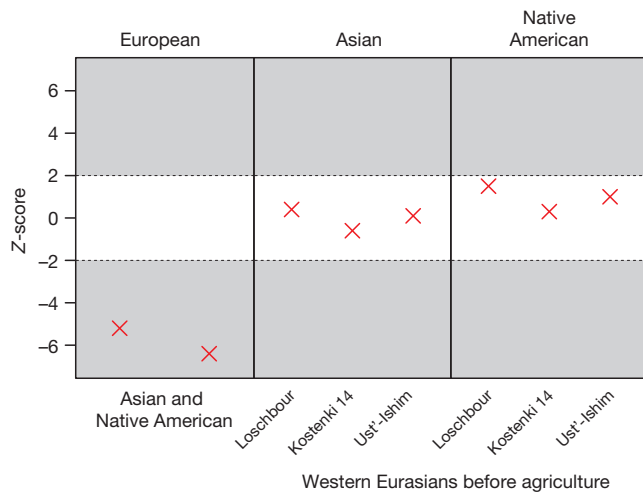


Figure 1 | Allele sharing between the Oase 1 individual and other genomes. Each point indicates the extent to which the Oase 1 genome shares alleles with one or other of a pair of genomes from different populations indicated above and below (see Extended Data Table 1 for numbers). Z-scores with an absolute value greater than 2 indicate an excess of allele sharing (grey).

among all mtDNA fragments to be 67% (95% confidence interval 65–69%). When we restrict to mtDNA fragments that carry terminal C-to-T substitutions, the contamination estimate is 4% (95% confidence interval of 2–9%) (Supplementary Note 1).

To isolate nuclear DNA from Oase 1, we used three sets of oligonucleotide probes that cover about two million sites that are single nucleotide polymorphisms (SNPs) in present-day humans and captured DNA molecules from the five libraries. Of the SNPs targeted, 51% ($n = 1,038,619$) were covered by at least one DNA fragment, and 13% ($n = 271,326$) were covered by at least one fragment with a terminal C-to-T substitution. To estimate nuclear DNA contamination, we tested whether Oase 1 DNA fragments with or without evidence of deamination share more alleles with present-day Europeans or with East Asians. We found that Europeans share significantly fewer alleles with Oase 1 fragments that are deaminated than with Oase 1 fragments that are not, consistent with European contamination of 17–30% (Supplementary Note 1). On the basis of these findings and those from mtDNA, we restricted all subsequent analyses to DNA fragments that carry terminal C-to-T substitutions. After doing this, we found that we captured targeted SNPs from the X and Y chromosomes at a similar rate, indicating that Oase 1 carried both an X and a Y chromosome and thus that he was male. The Y chromosome alleles belong to the F haplogroup, which is carried by most males in Eurasia today (Supplementary Note 2).

To determine the relationship of the Oase 1 individual to present-day populations, we first tested whether he shared more alleles with

particular present-day individuals from different populations using D-statistics, which provides a robust estimate of admixture almost regardless of how SNPs for analysis are chosen²⁷. We find that Oase 1 shared more alleles with present-day East Asians and Native Americans than with present-day Europeans, counter to what might naively be expected for an ancient individual from Europe (Fig. 1) ($5.2 \leq |Z| \leq 6.4$; Extended Data Table 1). However, it has been suggested that Europeans after the introduction of agriculture derive a part of their ancestry from a ‘basal Eurasian’ population that separated from the initial settlers of Europe and Asia before they split from each other²⁸. Therefore, we replaced present-day Europeans with Palaeolithic and Mesolithic European individuals in these analyses. We then find that the Oase 1 individual shares equally many alleles with these early Europeans as with present-day East Asians and Native Americans (Fig. 1) ($|Z| \leq 1.5$ in Extended Data Table 1). Restricting this analysis to transversion polymorphisms, which are not susceptible to errors induced by cytosine deamination, does not influence this result (Extended Data Table 2 and Supplementary Note 3). This suggests that the Oase 1 individual belonged to a population that did not contribute much, or not at all, to later Europeans. This contrasts, for example, with the ~36,000–39,000-year-old Kostenki 14 individual from western Russia, who was more closely related to later Europeans than to East Asians ($1.9 \leq |Z| \leq 13.7$; Extended Data Table 1)⁴⁶.

To assess whether the ancestors of the Oase 1 individual mixed with Neanderthals, we tested whether the Altai Neanderthal genome shares more alleles with the Oase 1 genome than with sub-Saharan Africans. We find this to be the case ($|Z| = 7.7$; Supplementary Note 4). We then asked whether the amount of Neanderthal ancestry in the Oase 1 genome is similar to that in present-day non-Africans. Surprisingly, the Neanderthal genome shares more alleles with the Oase 1 individual than it does with any present-day people in Eurasia that we tested, indicating that he carries more Neanderthal-like DNA than present-day people ($5.0 \leq |Z| \leq 8.2$; Extended Data Table 3). We also observe more Neanderthal-like alleles in the Oase 1 individual when we compare him to four early modern humans: an 8,000-year-old individual from Luxembourg, and three individuals from Russia who vary in age between 24,000 and 45,000 years ($3.6 \leq |Z| \leq 6.8$; Extended Data Table 3). Thus, the Oase 1 individual appears to have carried more Neanderthal-like DNA than any other modern human analysed to date. This observation cannot be explained by residual present-day human contamination among the DNA fragments that carry terminal C-to-T substitutions, because all modern humans studied to date carry less Neanderthal ancestry than the Oase 1 genome, and thus contamination would lower, rather than increase, the apparent Neanderthal ancestry.

We estimated the proportion of Neanderthal DNA in the Oase 1 genome using three different statistics^{7,29} (Supplementary Note 4). Although the results differ, they all yield point estimates between 6.0% and 9.4% (Table 1). For one of the statistics, none of the 90% confidence intervals for Neanderthal ancestry in the other modern

Table 1 | Estimated fraction of the Oase 1 genome that derives from Neanderthals

Sample	Statistic 1 $f_4(\text{Denisova, Altai; Mbuti, X})$ $f_4(\text{Denisova, Altai; Mbuti, Mezmaiskaya})$			Statistic 2 $1 - \frac{f_4(\text{Mbuti, Chimp; X, Denisova})}{f_4(\text{Mbuti, Chimp; Dinka, Denisova})}$			Statistic 3 $\frac{f_4(\text{X, Mbuti; Denisova, Chimp})}{f_4(\text{Altai, Mbuti; Denisova, Chimp})}$		
	Proportion	s.e.m.	90% CI	Proportion	s.e.m.	90% CI	Proportion	s.e.m.	90% CI
Oase 1	8.1%	2.0%	4.8–11.3%	9.4%	1.1%	7.5–11.3%	6.0%	2.0%	2.8–9.3%
Ust'-Ishim	3.6%	0.9%	2.2–5.0%	5.5%	0.7%	4.3–6.6%	0.4%	1.2%	0.0–2.5%
Kostenki 14	3.8%	1.0%	2.1–5.5%	2.9%	0.8%	1.6–4.2%	1.7%	1.3%	0.0–3.9%
MA1	1.2%	1.1%	0.0–3.0%	3.5%	0.8%	2.2–4.8%	2.3%	1.3%	0.1–4.5%
Loschbour	1.3%	0.9%	0.0–2.8%	3.9%	0.7%	2.7–5.1%	0.5%	1.2%	0.0–2.6%
La Braña	3.1%	1.0%	1.4–4.7%	1.9%	0.7%	0.7–3.1%	1.4%	1.2%	0.0–3.4%
Stuttgart	3.0%	0.9%	1.5–4.4%	2.5%	0.7%	1.3–3.7%	0.4%	1.2%	0.0–2.4%
Han	2.2%	0.9%	0.6–3.7%	2.2%	0.8%	1.0–3.5%	1.0%	1.2%	0.0–3.1%
Dai	2.6%	0.9%	1.1–4.0%	1.0%	0.8%	0.0–2.3%	0.7%	1.2%	0.0–2.6%
French	3.0%	0.9%	1.6–4.5%	3.0%	0.7%	1.8–4.2%	0.2%	1.2%	0.0–2.2%

CI, confidence interval; s.e.m., standard error of the mean; negative values are truncated to 0%.

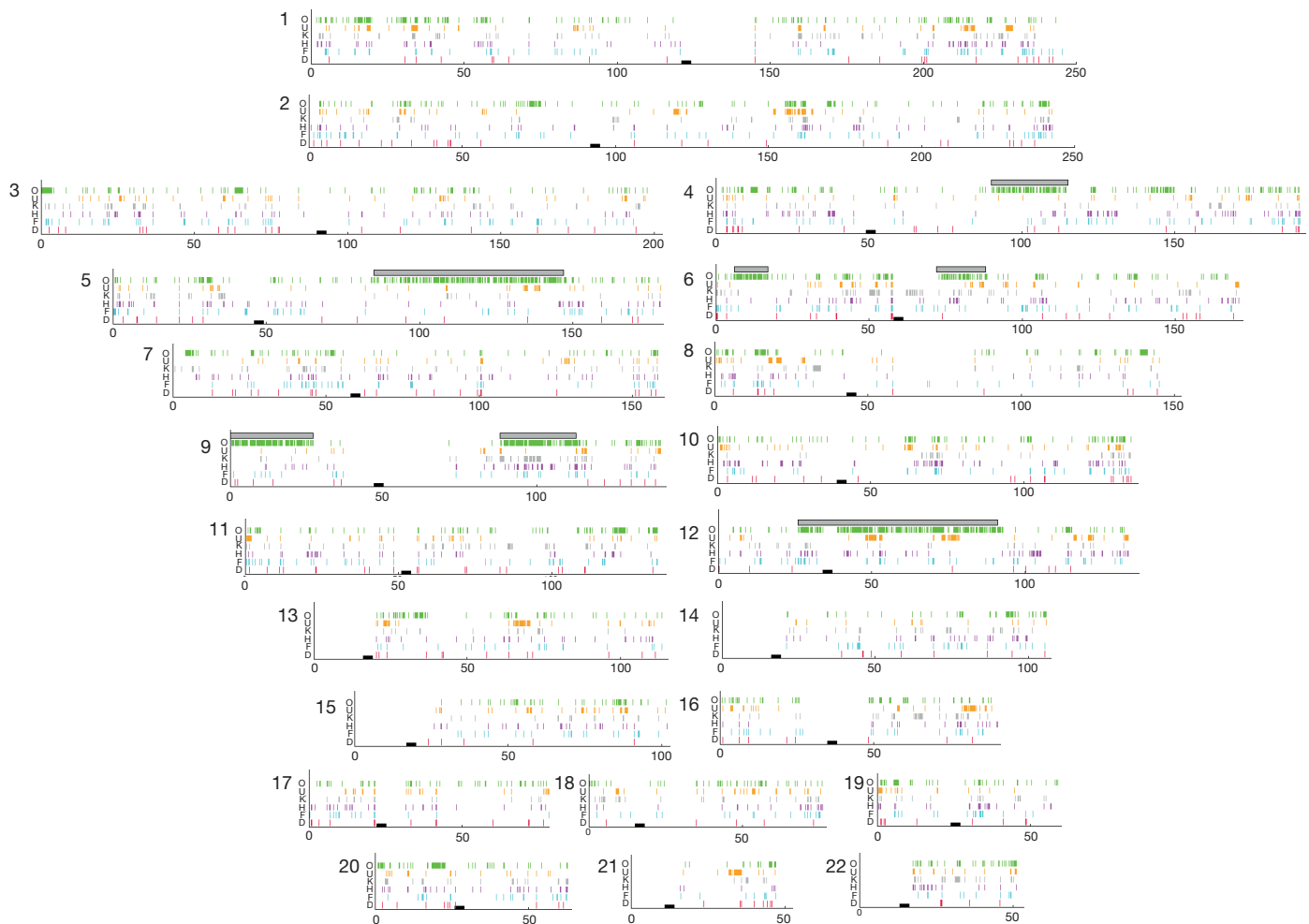


Figure 2 | Spatial distribution of alleles matching Neanderthals in modern humans. Coloured vertical lines indicate alleles shared with Neanderthals and no colour indicates alleles shared with the great majority of West Africans.

human samples overlap with the confidence interval in Oase 1. When we restrict analysis to transversion SNPs, the point estimates of Neanderthal ancestry are even higher (range of 8.4% to 11.3%) (Extended Data Table 4).

To study the spatial distribution of Neanderthal DNA across the Oase 1 genome, we designed capture probes for around 1.7 million nucleotide positions at which nearly all individuals in a sub-Saharan African population carry one allele whereas Neanderthal genomes carry a different allele. We used these probes to isolate DNA fragments from the Oase 1 individual. A total of 78,055 sites were covered by deaminated DNA fragments from the Oase 1 individual and were also covered by DNA fragments sequenced from the ~36,000–39,000-year-old Kostenki 14 individual from western Russia¹⁶, the ~43,000–47,000-year-old individual from Ust'-Ishim in Siberia¹⁵, and three present-day human genomes from China, France and Sudan (Supplementary Note 5). Because the Dinka from Sudan are thought to have little or no Neanderthal ancestry⁷, we subtracted the number of alleles that match the Neanderthals in the Dinka individual (485) from the number in the other genomes to estimate the number of alleles attributable to Neanderthal ancestry. The resulting numbers of putative Neanderthal alleles are 3,746 in the Oase 1 individual, 1,586 and 1,121 in the Ust'-Ishim and Kostenki 14 individuals, respectively, and 1,322 and 1,033 in the Chinese and the European individuals (Extended Data Table 5). Thus, the Neanderthal contribution to the Oase 1 genome appears to be between 2.3- and 3.6-fold larger than to the other genomes analysed. Assuming that the Neanderthal contri-

bution to the Oase 1 genome is 2% (ref. 7), this suggests that 7.3% of the Oase 1 genome is of Neanderthal origin. When the numbers of alleles matching the Neanderthal genome are compared per chromosome (Extended Data Table 5), the highest numbers are always observed for the Oase 1 genome, except in the case of chromosome 21, in which the Ust'-Ishim individual carries a large segment of likely Neanderthal ancestry.

We plotted the positions of Neanderthal-like alleles across the Oase 1 genome (Fig. 2). We detect three segments that are over 50 centimorgans (cM) in size, suggesting that the Neanderthal contribution to the Oase 1 individual occurred so recently in his family tree that chromosomal segments of Neanderthal origin had little time to break up due to recombination. To estimate the date of the most recent Neanderthal contribution to the Oase 1 genome, we studied the size spans of seven segments of the genome that appeared to be recently derived from Neanderthals. Their genetic lengths suggest that the Oase 1 individual had a Neanderthal ancestor as a fourth-, fifth- or sixth-degree relative (Supplementary Note 5). This would predict that an average of 1.6% to 6.3% of the Oase 1 genome derived from this recent Neanderthal ancestor. Visual inspection of the Oase 1 genome suggests that in addition to these seven segments, other smaller segments also carry Neanderthal-like alleles (Fig. 2). When we remove the seven longest segments, the estimate of Neanderthal ancestry in Oase 1 drops from 7.3% to 4.8%, which is still around twice the 2.0–2.9% estimated for the French, Han, Kostenki and Ust'-Ishim individuals in this remaining part of the genome. This additional Neanderthal ancestry

could reflect an older Neanderthal admixture into the ancestors of Oase 1, or that we failed to find all segments of recent Neanderthal ancestry.

The Oase 1 genome shows that mixture between modern humans and Neanderthals was not limited to the first ancestors of present-day people to leave Africa, or to people in the Near East; it occurred later as well and probably in Europe. The fact that the Oase 1 individual had a Neanderthal ancestor removed by only four to six generations allows this Neanderthal admixture to be dated to less than 200 years before the time he lived. However, the absence of a clear relationship of the Oase 1 individual to later modern humans in Europe suggests that he may have been a member of an initial early modern human population that interbred with Neanderthals but did not contribute much to later European populations. To better understand the interactions between early modern and Neanderthal populations, it will be important to study other specimens that, like Oase 1, have been suggested to carry morphological traits suggestive of admixture with Neanderthals³⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 February 2015; accepted 5 May 2015.

Published online 22 June 2015.

1. Higham, T. *et al.* The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512**, 306–309 (2014).
2. Trinkaus, E. *et al.* An early modern human from the Peștera cu Oase, Romania. *Proc. Natl Acad. Sci. USA* **100**, 11231–11236 (2003).
3. Smith, F. H. *et al.* The assimilation model, modern human origins in Europe, and the extinction of Neandertals. *Quat. Int.* **137**, 7–19 (2005).
4. Zilhão, J. Neandertals and moderns mixed, and it matters. *Evol. Anthropol.* **15**, 183–195 (2006).
5. Hublin, J.-J. The modern human colonization of western Eurasia: when and where? *Quat. Sci. Rev.* **118**, 194–210 (2014).
6. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
7. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
8. Sankararaman, S. *et al.* The date of interbreeding between Neandertals and modern humans. *PLoS Genet.* **8**, e1002947 (2012).
9. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
10. Wall, J. D. *et al.* Higher levels of Neanderthal ancestry in East Asians than in Europeans. *Genetics* **194**, 199–209 (2013).
11. Kim, B. Y. & Lohmueller, K. E. Selection and reduced population size cannot explain higher amounts of Neandertal ancestry in East Asian than in European human populations. *Am. J. Hum. Genet.* **96**, 454–461 (2015).
12. Vernot, B. & Akey, J. M. Complex history of admixture between modern humans and Neandertals. *Am. J. Hum. Genet.* **96**, 448–453 (2015).
13. d'Errico, F. *et al.* Neanderthal acculturation in Western Europe? A critical review of the evidence and its interpretation. *Curr. Anthropol.* **39** (suppl.), S1–S44 (1998).
14. Mellars, P. The impossible coincidence: a single-species model for the origins of modern human behavior in Europe. *Evol. Anthropol.* **14**, 12–27 (2005).
15. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
16. Seguin-Orlando, A. *et al.* Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
17. Rougier, H. *et al.* Peștera cu Oase 2 and the cranial morphology of early modern Europeans. *Proc. Natl Acad. Sci. USA* **104**, 1165–1170 (2007).
18. Constantin, S. *et al.* Reconstructing the evolution of cave systems as a key to understanding the taphonomy of fossil accumulations: the case of Urșilor Cave (Western Carpathians, Romania). *Quat. Int.* **339–340**, 25–40 (2014).
19. Ackermann, R. R. Phenotypic traits of primate hybrids: Recognizing admixture in the fossil record. *Evolutionary Anthropology: Issues, News, and Reviews* **19**, 258–270 (2010).
20. Frayer, D. in *Continuity or Replacement? Controversies in Homo sapiens Evolution* (eds Bräuer, G. & Smith, F. H.) (A.A. Balkema Publishers, 1992).
21. Trinkaus, E. *et al.* *Life and Death at the Peștera cu Oase. A Setting for Modern Human Emergence in Europe* (Oxford Univ. Press, 2013).
22. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
23. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
24. Sawyer, S. *et al.* Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* **7**, e34131 (2012).
25. Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403–406 (2014).
26. Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111**, 2229–2234 (2014).
27. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
28. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
29. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2014).
30. Trinkaus, E. European early modern humans and the fate of the Neandertals. *Proc. Natl Acad. Sci. USA* **104**, 7367–7372 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Trinkaus for facilitating the sampling of Oase 1, J. Krause for help with sampling and initial DNA analyses, A. Aximu for help with DNA hybridization captures and library preparation, and E. Trinkaus and L. Vigilant for critical reading of the manuscript. Q.F. is funded in part by the Chinese Academy of Sciences (XDA05130202) and the Special Foundation of the President of the Chinese Academy of Sciences (2015–2016). O.T.M. and S.C. are supported by the Romanian National Research Council through project PCCE 31/2010 (Karst Climate Archives). D.R. is supported by US National Science Foundation HOMINID grant BCS-1032255, US National Institutes of Health grant GM100233, and the Howard Hughes Medical Institute. The laboratory work was funded by the Presidential Innovation Fund of the Max Planck Society.

Author Contributions N.P., K.P., M.M., J.K., D.R. and S.P. supervised the study. S.C. and O.T.M. collected and analysed archaeological material. Q.F., M.H. and B.N. performed laboratory work. Q.F., M.H., S.M., P.S., N.P., N.R., I.L., B.V., K.P., J.K. and D.R. analysed data. Q.F., S.M., M.M. and D.R. designed capture probes. D.R. and S.P. wrote the manuscript with the help of all co-authors.

Author Information The aligned sequences have been deposited in the European Nucleotide Archive under accession number PRJEB8987. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.R. (reich@genetics.med.harvard.edu) or S.P. (paabo@eva.mpg.de).

METHODS

DNA extraction and library preparation. We used a dentistry drill to remove two samples of bone powder from an area where a larger sample had previously been removed for carbon dating². We prepared two extracts (E1406, E1843) from 25 mg and 10 mg of bone powder, respectively, as described³¹. We produced five libraries from the two extracts using a single-stranded library protocol^{19,32} (Extended Data Table 6). We treated one library from each extract (A5227, A5252) with *E. coli* uracil-DNA glycosylase (UDG) and endonuclease VIII to remove deaminated cytosine residues from the interior parts of molecules³³. We amplified all libraries by PCR for 35 cycles using AccuPrime Pfx DNA polymerase (Life Technologies)³⁴ and primers carrying library-specific indexes³⁵. We determined library concentrations using a NanoDrop 2000 spectrophotometer.

Sequencing and DNA capture. We shotgun sequenced the UDG-treated libraries A5252 and A5227 and found that they contained 0.06% and 0.18% human DNA, respectively. We used hybridization to oligonucleotide probes to enrich the libraries for subsets of the nuclear genome containing panels of known SNPs as described²³, except that each SNP was targeted by four 52-nucleotide probes: two immediately flanking the SNP on both sides, and two centred on the SNP containing one or the other alternate allele, respectively. We used four panels of probes.

Panel 1 “390k”: 394,577 SNPs, about 90% of which are on the Affymetrix Human Origins array²⁷. See ref. 36 for SNPs and probes.

Panel 2 “840k”: 842,630 SNPs constituting the rest of the SNPs on the Human Origins array, all SNPs on the Illumina 610-Quad array, all SNPs on the Affymetrix 50k array, and smaller numbers of SNPs chosen for other purposes. See Supplementary Data 1.

Panel 3 “1000k”: 997,780 SNPs comprising all transversion polymorphisms seen in two Yoruba males from Nigeria sequenced to high coverage and transversion polymorphisms seen in the Altai Neanderthal genome. The design was restricted to SNPs that passed strict quality filters in the Neanderthal genome (Map35_99%)⁷, and had chimpanzee alleles available. Probes were designed from chimpanzee flanking sequences. See Supplementary Data 2.

Panel 4 “Archaic”: This panel contains SNPs where the West-African Yoruba population carry a high frequency of one allele while at least one archaic individual carries an alternative allele. To determine Yoruba allele frequencies, we examined data from all Yoruba individuals from the 1000 Genomes Project³⁷ covered by at least three sequences passing filters. At these sites we called majority alleles (drawing a random allele in the case of equal numbers of reads supporting both alleles). We furthermore restricted the analysis to sites at which ≥ 24 Yoruba individuals as well as the Altai Neanderthal and Denisovan genomes had allele calls (Map35_50% filter⁷). We then selected sites at which at most one alternative allele is seen among the Yoruba while at least one of four archaic genomes (Denisovan; Altai, Vindija and Mezmaiskaya Neanderthals) carry the alternative allele. Ancestral states were taken from the inferred ancestor of humans and chimpanzees (Ensembl Compara v.64)^{38,39}. We used the following classes of sites. Class 1: 297,894 SNPs where Yoruba is derived and at least one ancestral allele is seen in the Altai, Vindija, Mezmaiskaya or Denisova genomes. Class 2: sites where Yoruba alleles are all or nearly all ancestral and derived alleles are seen in archaic genomes. Since such derived alleles often arise due to errors in an archaic genome, we restricted this class to the following three cases: (1) 1,321,774 SNPs where the high-coverage Altai Neanderthal and/or Denisova genomes are homozygous derived; (2) 523,041 SNPs where the Altai and/or Denisova genomes are heterozygous but are not C-to-T or G-to-A substitutions relative to the ancestral allele; and (3) 30,735 SNPs that are homozygous ancestral in Altai and/or Denisova and at least one copy of the derived allele is observed in the Mezmaiskaya or Vindija Neanderthal genomes, and the derived allele represents a transversion that is also seen in the Simons Genome Diversity Panel (<https://www.simonsfoundation.org/life-sciences/simons-genome-diversity-project/>). After eliminating SNPs where capture probes covered ambiguous bases in the human (hg19) and chimpanzee (panTro2) genomes or overlapped for less than 35 nucleotides with mappable regions (Map35_50%)⁷, this left us with a set of 1,749,385 SNPs (see Supplementary Data 3).

Sequencing of capture products and data processing. We sequenced capture products using 2×75 bp reads on an Illumina HiSeq2500 or an Illumina NextSeq500. We de-multiplexed the reads allowing one mismatch in each of the

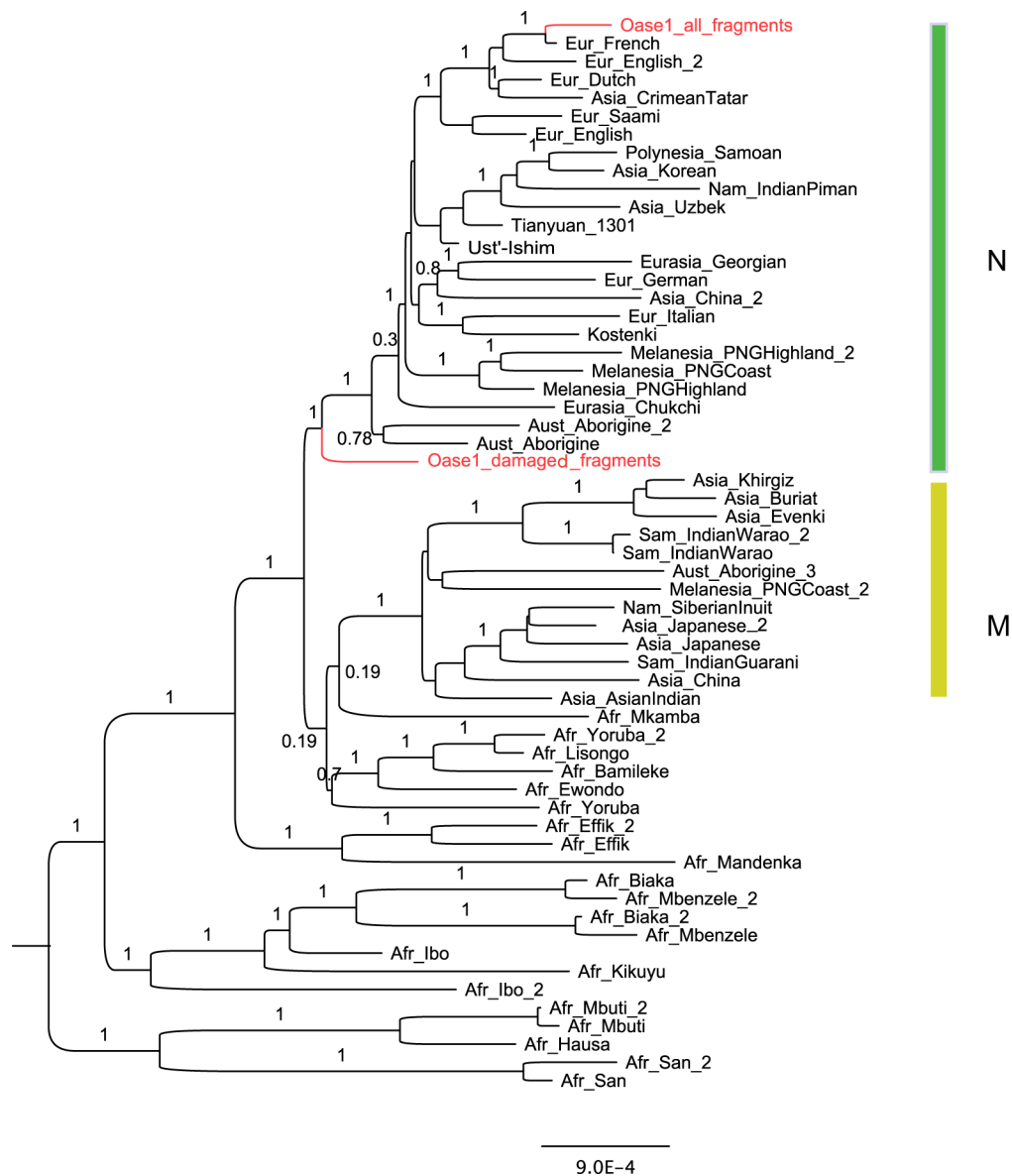
two indices (Extended Data Table 6), and merged paired reads into sequenced fragments requiring an overlap of at least 15 bp (allowing one mismatch) using a modified form of SeqPrep (<https://github.com/jstjohn/SeqPrep>). We used the bases with the higher quality (and score) to represent the overlap region. After removing adapters, we mapped merged fragments to hg19 using BWA (v.0.6.1) using the ‘samse’ command. We identified duplicated fragments on the basis of sharing the same orientation and end positions, in which case we kept the fragment with the highest quality (Extended Data Table 7).

To focus on putatively deaminated fragments we used fragments with C-to-T substitutions relative to the hg19 human genome reference sequence in the first 5’ or last two 3’ bases for the UDG-treated libraries, and to fragments with C-to-T substitutions relative to hg19 in the terminal three bases at either end of fragments from non-UDG-treated libraries (Supplementary Note 1 and Extended Data Table 8).

Merging the Oase 1 data with genome sequences. At each SNP covered at least once in Oase 1, we selected the majority allele (in case of a tie, we picked a random allele). We then merged the Oase 1 data with 25 genomes of present-day humans sequenced to 24–42 \times coverage⁷: the Altai Neanderthal⁷, the Siberian Denisovan⁹, a $\sim 45,000$ -year-old modern human from Ust’-Ishim in Siberia¹⁵, an $\sim 8,000$ -year-old Mesolithic individual from Loschbour Cave, Luxembourg²⁸, and a $\sim 7,000$ -year-old early farmer from Stuttgart, Germany²⁸ (Extended Data Table 9). All the genotype calls for the five deeply sequenced ancient genomes were performed in the same way. We restricted analyses to sites with a minimum root-mean-square mapping quality (MAPQ) of 30 in the 30 genomes. We added lower coverage shotgun data from the $\sim 36,000$ -year-old Kostenki 14 from Russia¹⁶, the $\sim 24,000$ -year-old Mal’ta Siberian individual from Russia⁴⁰, an 8,000-year-old Mesolithic individual from La Braña Cave, Spain⁴¹, a Neanderthal from Mezmaiskaya in Russia⁷, and a pool of three Neanderthals from Vindija Cave in Croatia⁶. For these samples, we restricted to fragments with a map quality of MAPQ ≥ 37 to match the filter for the low-coverage Oase 1 data (Extended Data Table 9).

Population genetic analyses. To determine the relationship of Oase 1 to other modern humans, we used *D*-statistics to evaluate whether sets of four tested samples are consistent with being related to one another according to an unrooted tree²⁷ (Supplementary Note 3). We used *D*-statistics and *f_d*-statistic ratios²⁷ to test both whether there is excess archaic ancestry in Oase 1 compared with other modern humans, and to estimate proportions of Neanderthal ancestry²⁷ (Supplementary Note 4). We studied the genomic distribution of alleles that are likely to derive from Neanderthals in the sense of being shared with Neanderthal but either absent or at very low frequency in West Africans. We used the spatial distribution of these sites to identify stretches of likely Neanderthal ancestry in several individuals including Oase 1. We also used these data to estimate the number of generations since the most recent Neanderthal ancestor of Oase 1 (Supplementary Note 5).

- Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
- Gansauge, M. T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**, 737–748 (2013).
- Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neanderthal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).
- Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87–94 (2012).
- Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
- Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
- The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- Paten, B., Herrero, J., Beal, K., Fitzgerald, S. & Birney, E. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res.* **18**, 1814–1828 (2008).
- Paten, B. *et al.* Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.* **18**, 1829–1843 (2008).
- Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
- Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).



Extended Data Figure 1 | Mitochondrial DNA tree for Oase 1 and other modern humans. The consensus sequences for all Oase 1 fragments and for deaminated fragments are shown. The tree is rooted with a Neanderthal mtDNA (Vindija33.25).

Extended Data Table 1 | Allele sharing between early modern humans and other humans

<i>Non-African₁</i>	<i>Non-African₂</i>	Oase 1		Ust'-Ishim		Kostenki 14	
		D	Z	D	Z	D	Z
Oase 1	Ust'-Ishim					-0.0033	-3.8
Oase 1	Kostenki 14			-0.0037	-4.1		
Oase 1	MA1			-0.0032	-3.5	-0.0092	-9.8
Oase 1	Loschbour			-0.0032	-3.9	-0.0101	-12.2
Oase 1	East Asian			-0.0027	-3.8	-0.0011	-1.6
Oase 1	Native American			-0.0030	-4.1	-0.0039	-5.5
Ust'-Ishim	Kostenki 14	-0.0005	-0.6				
Ust'-Ishim	MA1	-0.0007	-0.8			-0.0059	-6.4
Ust'-Ishim	Loschbour	0.0002	0.3			-0.0068	-8.5
Ust'-Ishim	East Asian	0.0000	-0.1			0.0022	3.3
Ust'-Ishim	Native American	-0.0007	-1.0			-0.0006	-0.8
Kostenki 14	MA1	-0.0004	-0.6	0.0003	0.4		
Kostenki 14	Loschbour	0.0007	1.0	0.0006	0.8		
Kostenki 14	East Asian	0.0004	0.6	0.0011	1.6		
Kostenki 14	Native American	-0.0002	-0.3	0.0008	1.1		
MA1	Loschbour	0.0012	1.7	0.0005	0.7	-0.0012	-1.5
MA1	East Asian	0.0008	1.2	0.0007	1.1	0.0079	10.6
MA1	Native American	0.0001	0.1	0.0004	0.6	0.0051	7.0
Loschbour	East Asian	-0.0002	-0.4	0.0005	0.9	0.0090	13.7
Loschbour	Native American	-0.0009	-1.5	0.0002	0.3	0.0062	9.0
East Asian	Native American	-0.0006	-1.6	-0.0003	-0.8	-0.0028	-6.6
European	Oase 1			0.0004	0.6	0.0049	7.3
European	Ust'-Ishim	-0.0023	-3.5			0.0016	2.4
European	Kostenki 14	-0.0028	-4.7	-0.0033	-5.1		
European	MA1	-0.0033	-5.4	-0.0031	-5.1	-0.0041	-6.0
European	Loschbour	-0.0021	-4.5	-0.0027	-5.7	-0.0052	-9.1
European	East Asian	-0.0024	-5.2	-0.0022	-5.3	0.0039	9.2
European	Native American	-0.0030	-6.4	-0.0025	-5.9	0.0010	2.2
European	Stuttgart	-0.0007	-1.5	-0.0001	-0.2	-0.0002	-0.3
Stuttgart	Oase 1			0.0005	0.6	0.0051	6.7
Stuttgart	Ust'-Ishim	-0.0017	-2.3			0.0018	2.3
Stuttgart	Kostenki 14	-0.0021	-3.2	-0.0032	-4.6		
Stuttgart	MA1	-0.0027	-3.9	-0.0029	-4.2	-0.0041	-5.0
Stuttgart	Loschbour	-0.0015	-2.4	-0.0027	-4.6	-0.0050	-7.5
Stuttgart	East Asian	-0.0017	-2.9	-0.0022	-3.8	0.0040	6.8
Stuttgart	Native American	-0.0024	-3.9	-0.0025	-4.4	0.0012	1.9

We compute $D(\text{Non-African}_1, \text{Non-African}_2; \text{Early Modern Human, African})$ to test whether an early modern human (Oase 1, Ust'-Ishim, or Kostenki 14) shares more alleles with Non-African_1 (in which case the statistic is positive) or Non-African_2 (negative). We use a pool of six sub-Saharan African genomes (2 Mbuti, 2 Yoruba, 2 Dinka) as an outgroup; a pool of four genomes (2 French, 2 Sardinians) to represent Europeans; a pool of four genomes (2 Han, 2 Dai) to represent East Asians; and a pool of three genomes (2 Karitiana, 1 Mixe) to represent Native Americans. Results are based on 242,122 transition and transversion SNPs covered by at least one deaminated fragment in Oase 1, and covered in all other samples, although not necessarily MA1. For analyses involving MA1, a subset of 176,569 transversion SNPs was analysed.

Extended Data Table 2 | Allele sharing between early modern humans and other humans (transversions only)

<i>Non-African₁</i>	<i>Non-African₂</i>	Oase 1		Ust'-Ishim		Kostenki 14	
		D	Z	D	Z	D	Z
Oase 1	Ust'-Ishim					-0.0019	-2.1
Oase 1	Kostenki 14			-0.0031	-3.3		
Oase 1	MA1			-0.0026	-2.9	-0.0071	-6.5
Oase 1	Loschbour			-0.0023	-2.6	-0.0081	-8.8
Oase 1	East Asian			-0.0013	-1.9	0.0007	1.0
Oase 1	Native American			-0.0019	-2.7	-0.0018	-2.3
Ust'-Ishim	Kostenki 14	-0.0012	-1.4				
Ust'-Ishim	MA1	-0.0006	-0.7			-0.0050	-5.1
Ust'-Ishim	Loschbour	0.0003	0.4			-0.0062	-7.1
Ust'-Ishim	East Asian	0.0005	0.7			0.0026	3.8
Ust'-Ishim	Native American	-0.0003	-0.4			0.0001	0.1
Kostenki 14	MA1	0.0001	0.1	0.0002	0.3		
Kostenki 14	Loschbour	0.0015	2.0	0.0008	1.1		
Kostenki 14	East Asian	0.0017	2.3	0.0017	2.5		
Kostenki 14	Native American	0.0009	1.2	0.0012	1.6		
MA1	Loschbour	0.0019	2.2	0.0010	1.3	-0.0013	-1.3
MA1	East Asian	0.0011	1.4	0.0013	1.9	0.0075	8.5
MA1	Native American	0.0006	0.7	0.0007	1.1	0.0051	6.0
Loschbour	East Asian	0.0001	0.2	0.0009	1.5	0.0088	12.3
Loschbour	Native American	-0.0006	-0.9	0.0004	0.6	0.0063	8.4
East Asian	Native American	-0.0008	-1.7	-0.0006	-1.3	-0.0025	-5.3
European	Oase 1			-0.0005	-0.7	0.0029	3.9
European	Ust'-Ishim	-0.0023	-3.3			0.0010	1.4
European	Kostenki 14	-0.0035	-5.1	-0.0035	-5.2		
European	MA1	-0.0033	-4.5	-0.0033	-5.2	-0.0038	-4.8
European	Loschbour	-0.0020	-3.6	-0.0027	-5.1	-0.0052	-8.4
European	East Asian	-0.0018	-3.6	-0.0018	-4.0	0.0036	7.8
European	Native American	-0.0026	-4.8	-0.0023	-5.2	0.0011	2.1
European	Stuttgart	-0.0009	-1.7	-0.0010	-2.2	-0.0012	-2.3
Stuttgart	Oase 1			0.0005	0.7	0.0041	4.7
Stuttgart	Ust'-Ishim	-0.0014	-1.8			0.0022	2.6
Stuttgart	Kostenki 14	-0.0026	-3.3	-0.0025	-3.5		
Stuttgart	MA1	-0.0026	-3.1	-0.0023	-3.2	-0.0031	-3.4
Stuttgart	Loschbour	-0.0011	-1.6	-0.0017	-2.8	-0.0040	-5.2
Stuttgart	East Asian	-0.0010	-1.4	-0.0008	-1.3	0.0048	7.2
Stuttgart	Native American	-0.0017	-2.4	-0.0013	-2.2	0.0023	3.4

We compute $D(\text{Non-African}_1, \text{Non-African}_2; \text{Early Modern Human, African})$, to test whether an early modern human (Oase 1, Ust'-Ishim or Kostenki 14) shares more alleles with Non-African₁ (in which case the statistic is positive) or Non-African₂ (negative). We use a pool of six sub-Saharan African genomes (2 Mbuti, 2 Yoruba, 2 Dinka) as an outgroup; a pool of four genomes (2 French, 2 Sardinians) to represent Europeans; a pool of four genomes (2 Han, 2 Dai) to represent East Asians; and a pool of three genomes (2 Karitiana, 1 Mixe) to represent Native Americans. Statistics are as in Extended Data Table 1 but are based on 106,004 transversion SNPs covered by at least one deaminated fragment in Oase 1 and that also have coverage for all other samples, although not necessarily MA1. For analyses involving MA1, a subset of 76,715 transversion SNPs is analysed.

Extended Data Table 3 | Testing whether archaic genomes share more alleles with Oase 1 than with other modern humans

Test	Sites	Archaic = Altai				Archaic = Denisovan			
		Chimp		Mbuti		Chimp		Mbuti	
		D	Z	D	Z	D	Z	D	Z
Han	115,300	-0.0036	-5.1	-0.0071	-7.6	-0.0014	-2.2	-0.0049	-6.3
Dai	115,300	-0.0035	-5.0	-0.0077	-8.2	-0.0013	-2.1	-0.0056	-7.0
Karitiana	115,300	-0.0032	-4.3	-0.0063	-6.9	-0.0008	-1.3	-0.0040	-5.3
French	115,300	-0.0049	-6.9	-0.0074	-8.2	-0.0021	-3.4	-0.0047	-6.2
Sardinian	115,300	-0.0038	-5.1	-0.0071	-7.8	-0.0016	-2.5	-0.0050	-6.5
Papuan	115,300	-0.0026	-3.6	-0.0051	-5.4	0.0009	1.5	-0.0016	-2.1
Ust'-Ishim	115,100	-0.0026	-3.6	-0.0052	-5.5	-0.0009	-1.5	-0.0035	-4.4
Kostenki14	108,100	-0.0032	-4.1	-0.0059	-6.0	-0.0017	-2.4	-0.0044	-5.3
MA1	83,200	-0.0031	-3.6	-0.0050	-4.7	-0.0007	-0.9	-0.0028	-2.8
Loschbour	114,300	-0.0043	-5.7	-0.0066	-6.8	-0.0019	-2.9	-0.0043	-5.3
LaBrana	111,000	-0.0033	-4.2	-0.0072	-7.3	-0.0008	-1.2	-0.0047	-5.4
Stuttgart	114,000	-0.0037	-5.1	-0.0066	-7.1	-0.0013	-2.1	-0.0042	-5.6

The statistic $D(\text{Test}, \text{Oase 1}; \text{Archaic}, \text{Outgroup})$ is negative if the archaic genomes share more alleles with Oase 1 than with a test sample. The outgroups are either chimpanzee or a sub-Saharan African (Mbuti).

Extended Data Table 4 | Estimated fraction of the Oase 1 genome that derives from Neanderthals

Sample	$\frac{f_4(\text{Denisova,Altai;Mbuti,X})}{f_4(\text{Denisova,Altai;Mbuti,Mezmaiskaya})}$			$1 - \frac{f_4(\text{Mbuti,Chimp;X,Denisova})}{f_4(\text{Mbuti,Chimp;Dinka,Denisova})}$			$\frac{f_4(\text{X,Mbuti;Denisova,Chimp})}{f_4(\text{Altai,Mbuti;Denisova,Chimp})}$		
	Prop.	S.E.	90% CI	Prop.	S.E.	90% CI	Prop.	S.E.	90% CI
Oase 1	11.3%	2.8%	6.7%-16%	10.9%	1.6%	8.3%-13.6%	8.4%	2.7%	4.0%-12.9%
Ust'-Ishim	2.9%	1.2%	1.0%-4.9%	6.0%	0.8%	4.7%-7.4%	4.2%	1.5%	1.8%-6.6%
Kostenki 14	3.0%	1.4%	0.7%-5.3%	3.0%	0.9%	1.6%-4.5%	6.2%	1.6%	3.6%-8.7%
MA1	1.5%	1.5%	0.0%-4.0%	3.6%	1.0%	1.9%-5.2%	5.5%	1.6%	2.8%-8.2%
Loschbour	1.1%	1.2%	0.0%-3.1%	4.8%	0.9%	3.3%-6.2%	3.6%	1.5%	1.2%-6.1%
LaBrana	3.7%	1.3%	1.4%-5.9%	2.4%	0.9%	0.9%-3.8%	4.8%	1.5%	2.4%-7.2%
Stuttgart	2.8%	1.2%	0.8%-4.8%	3.4%	0.9%	2.0%-4.9%	3.8%	1.5%	1.4%-6.2%
Han	1.0%	1.3%	0.0%-3.1%	2.8%	0.9%	1.3%-4.2%	3.6%	1.5%	1.2%-6.1%
Dai	2.1%	1.2%	0.2%-4.0%	1.3%	0.9%	0.0%-2.8%	3.8%	1.5%	1.4%-6.2%
French	1.6%	1.2%	0.0%-3.5%	3.3%	0.9%	1.9%-4.7%	2.7%	1.5%	0.3%-5.2%
Sardinian	2.7%	1.2%	0.8%-4.7%	2.3%	0.9%	0.8%-3.7%	3.7%	1.4%	1.3%-6.1%

Estimates are as in Table 1 but restrict to transversions. Present-day human genomes are from a data set reported previously⁷.

Extended Data Table 5 | Counts of putative Neanderthal alleles in six modern humans

Chr	Sites	Neanderthal allele counts						Neanderthal ancestry				
		Oase 1	Ust'-Ishim	Kostenki 14	Han	French	Dinka	Oase 1	Ust'-Ishim	Kostenki 14	Han	French
1	6740	323	196	148	129	117	25	6.70%	3.84%	2.77%	2.34%	2.07%
2	7112	294	145	121	188	199	29	5.65%	2.47%	1.96%	3.39%	3.62%
3	5417	177	102	96	74	98	28	4.17%	2.07%	1.90%	1.29%	1.96%
4	4495	359	86	63	141	96	42	10.69%	1.48%	0.71%	3.34%	1.82%
5	4330	446	108	66	103	95	23	14.80%	2.97%	1.50%	2.80%	2.52%
6	4549	324	155	167	142	138	73	8.36%	2.73%	3.13%	2.30%	2.16%
7	4422	147	68	65	102	72	34	3.87%	1.16%	1.06%	2.33%	1.30%
8	4322	131	132	72	35	38	14	4.10%	4.14%	2.03%	0.74%	0.84%
9	3107	500	69	120	118	49	15	23.65%	2.63%	5.12%	5.02%	1.66%
10	4009	147	139	67	131	86	22	4.72%	4.42%	1.70%	4.12%	2.42%
11	4193	153	93	88	81	73	26	4.59%	2.42%	2.24%	1.99%	1.70%
12	3456	456	160	54	125	93	10	19.55%	6.58%	1.93%	5.04%	3.64%
13	2457	96	81	33	54	30	18	4.81%	3.89%	0.93%	2.22%	0.74%
14	2390	85	27	52	50	52	13	4.56%	0.89%	2.47%	2.35%	2.47%
15	2327	73	78	47	38	32	5	4.43%	4.75%	2.73%	2.15%	1.76%
16	3139	90	121	68	43	39	8	3.96%	5.45%	2.90%	1.69%	1.50%
17	2543	72	89	37	85	75	56	0.95%	1.97%	-1.13%	1.73%	1.13%
18	2305	57	58	59	27	29	5	3.42%	3.48%	3.55%	1.45%	1.58%
19	1769	79	49	33	43	35	12	5.74%	3.17%	1.80%	2.66%	1.97%
20	2492	107	29	62	56	43	12	5.78%	1.03%	3.04%	2.68%	1.88%
21	1026	36	53	22	8	11	10	3.84%	6.35%	1.77%	-0.30%	0.15%
22	1455	79	33	66	34	18	5	7.71%	2.92%	6.35%	3.02%	1.35%
All	78055	4231	2071	1606	1807	1518	485	7.27%	3.08%	2.18%	2.57%	"2%"
Subtract Dinka		3746	1586	1121	1322	1033	0					

The analysis is based on 78,055 sites covered by at least one deaminated fragment in Oase 1. To convert the counts to estimates of ancestry, we subtract the Dinka count as an estimate of the false positive rate and divide by the number of sites covered (as indicated for the whole genome on the bottom). This gives the rate of alleles per screened site on this chromosome for this individual. We then multiply this quantity by 2%/1.32% to recalibrate the 1.32% seen genome-wide in the French to an assumed 2% genome-wide Neanderthal ancestry in the French⁷.

Extended Data Table 6 | Ancient DNA libraries made from the Oase 1 mandible

Metainformation						Sequencing results			All fragments			Deaminated fragments		
Library	Ex-tract	UDG treatment	Index 1	Index 2	Extra-ct used (μl)	Sequences going into alignment	Sequences ≥35bp mapped	After dup. removal	Cov-er-age	% C→T 5' end	% C→T 3' end	Cov-er-age	% C→T 5' end	% C→T 3' end
A5227	E1406	Yes	ACTTGCG	AACTCCG	8	206,982	118,976	34,486	112	8	19	5	19	36
A5252	E1843	Yes	GTAAGCC	TTGAAGT	40	74,384	46,394	31,368	114	7	25	5	18	55
A9032	E1406	No	ATAACGT	ACTATCA	6	9,321,903	5,904,210	51,810	178	20	21	12	31	39
A9033	E1406	No	AATAGGA	ACCAACT	6	7,932,271	4,816,314	55,878	193	21	20	13	36	38
A9034	E1406	No	ATCACGA	AACTCCG	6	10,422,467	6,861,634	59,883	207	20	20	14	35	38
						27,958,007	17,747,528	233,425	803	17	21	49	30	39

Extended Data Table 7 | Sequencing metrics on the five libraries for the four capture probe panels

Library	Panel	No. target SNPs	Fragments going into alignment	Fragments mapped to genome	Fragments on target after dup. removal and MAPQ37 filter	% SNPs hit at least once	Average coverage on SNPs
A9032	390k	393,577	10,849,144	2,235,955	133,564	26.5%	0.34
A9033	390k	393,577	17,159,085	2,808,704	73,824	15.9%	0.19
A9034	390k	393,577	16,902,935	3,256,438	142,520	27.7%	0.36
A5227	390k	393,577	63,441,719	22,124,247	195,161	36.0%	0.5
A5252	390k	393,577	60,181,844	14,278,978	180,626	33.3%	0.46
All 5	390k	393,577	168,534,727	44,704,322	724,653	73.0%	1.84
A9032	840k	842,630	25,105,625	3,801,435	178,015	17.6%	0.21
A9033	840k	842,630	29,196,969	4,655,434	183,093	17.9%	0.22
A9034	840k	842,630	35,780,652	5,968,851	200,767	19.3%	0.24
A5227	840k	842,630	28,209,496	4,276,439	152,411	15.3%	0.18
A5252	840k	842,630	20,286,540	1,630,343	106,943	11.2%	0.13
All 5	840k	842,630	138,579,282	20,332,502	818,648	51.7%	0.97
A9032	1000k	997,780	26,088,835	2,964,094	159,162	13.5%	0.16
A9033	1000k	997,780	26,641,358	4,490,372	158,614	13.3%	0.16
A9034	1000k	997,780	28,795,043	4,985,140	154,177	13.0%	0.15
A5227	1000k	997,780	25,848,311	4,395,413	71,537	6.4%	0.07
A5252	1000k	997,780	25,691,323	2,254,636	53,932	5.0%	0.05
All 5	1000k	997,780	133,064,870	19,089,655	596,107	36.1%	0.6
A9032	Archaic	1,749,385	19,329,832	2,086,208	205,095	10.0%	0.12
A9033	Archaic	1,749,385	24,629,023	2,768,355	237,818	11.4%	0.14
A9034	Archaic	1,749,385	31,200,466	3,783,805	257,351	12.2%	0.15
A5227	Archaic	1,749,385	27,659,125	3,606,375	195,356	9.6%	0.11
A5252	Archaic	1,749,385	31,472,143	2,435,080	136,637	6.8%	0.08
All 5	Archaic	1,749,385	134,290,589	14,679,823	1,022,046	34.6%	0.58
A9032	Combined	3,801,245	81,373,436	11,087,692	719,146	15.5%	0.19
A9033	Combined	3,801,245	97,626,435	14,722,865	698,890	15.1%	0.18
A9034	Combined	3,801,245	112,679,096	17,994,234	806,589	17.0%	0.21
A5227	Combined	3,801,245	145,158,651	34,402,474	666,195	14.2%	0.18
A5252	Combined	3,801,245	137,631,850	20,599,037	531,873	11.4%	0.14
All 5	Combined	3,801,245	574,469,468	98,806,302	3,406,685	45.5%	0.90

Extended Data Table 8 | Effect of filtering on amount of nuclear data available

Panel	Target SNPs	All fragments			Deaminated fragments only		
		No. SNPs hit ≥1×	% SNPs hit ≥1×	Average coverage	No. SNPs hit ≥1×	% SNPs hit ≥1×	Average coverage
Panels 1-3	2,051,902	1,038,619	50.6%	1.03	271,326	13.2%	0.16
Panel 4 subset*	954,849	361,681	37.9%	0.69	87,803	9.2%	0.11
Panels 1-4	3,801,245	1,685,891	44.4%	0.85	426,027	11.2%	0.13

Note that numbers differ from Extended Data Table 7 because only sites with base quality ≥20 were used.
* The Panel 4 subset excludes the sites where only the Denisovan genome differs from the African panel.

Extended Data Table 9 | Genomes merged with the Oase 1 data

Sample ID	Human	Data type	Mean	UDG-treated
Oase1	Modern	Low coverage	Capture	Mix of library types
Vindija	Archaic	Low coverage	1.3	No
Mezmaiskaya	Archaic	Low coverage	0.5	Yes
Altai	Archaic	High coverage	52	Yes
Denisova	Archaic	High coverage	31	Yes
Kostenki14	Modern	Low coverage	2.4	Mix of library types
MA1	Modern	Low coverage	1	No
LaBrana	Modern	Low coverage	3.4	No
Loschbour	Modern	High coverage	22	Yes
Stuttgart	Modern	High coverage	19	Yes
Ust'-Ishim	Modern	High coverage	42	Yes
Dinka _A	Modern	High coverage	28	..
French _A	Modern	High coverage	27	..
Papuan _A	Modern	High coverage	26	..
Sardinian _A	Modern	High coverage	25	..
Han _A	Modern	High coverage	28	..
Yoruba _A	Modern	High coverage	32	..
Karitiana _A	Modern	High coverage	26	..
San _A	Modern	High coverage	33	..
Mandenka _A	Modern	High coverage	25	..
Dai _A	Modern	High coverage	28	..
Mbuti _A	Modern	High coverage	24	..
Dai _B	Modern	High coverage	37	..
French _B	Modern	High coverage	42	..
Han _B	Modern	High coverage	35	..
Mandenka _B	Modern	High coverage	37	..
Mbuti _B	Modern	High coverage	37	..
Papuan _B	Modern	High coverage	42	..
San _B	Modern	High coverage	38	..
Sardinian _B	Modern	High coverage	38	..
Yoruba _B	Modern	High coverage	39	..
Karitiana _B	Modern	High coverage	35	..
Mixe _B	Modern	High coverage	42	..
Australian _{B1}	Modern	High coverage	42	..
Australian _{B2}	Modern	High coverage	37	..
Dinka _B	Modern	High coverage	35	..

For the 25 present-day humans, individuals ending with a subscript 'A' are from 'Panel A' reported in ref. 9 and individuals with a subscript 'B' are from 'Panel B' reported in ref. 7. Unless otherwise specified, we used Panel B individuals.

The octopus genome and the evolution of cephalopod neural and morphological novelties

Caroline B. Albertin^{1*}, Oleg Simakov^{2,3*}, Therese Mitros⁴, Z. Yan Wang⁵, Judit R. Pungor⁵, Eric Edsinger-Gonzales^{2,4}, Sydney Brenner², Clifton W. Ragsdale^{1,5} & Daniel S. Rokhsar^{2,4,6}

Coleoid cephalopods (octopus, squid and cuttlefish) are active, resourceful predators with a rich behavioural repertoire¹. They have the largest nervous systems among the invertebrates² and present other striking morphological innovations including camera-like eyes, prehensile arms, a highly derived early embryogenesis and a remarkably sophisticated adaptive colouration system^{1,3}. To investigate the molecular bases of cephalopod brain and body innovations, we sequenced the genome and multiple transcriptomes of the California two-spot octopus, *Octopus bimaculoides*. We found no evidence for hypothesized whole-genome duplications in the octopus lineage^{4–6}. The core developmental and neuronal gene repertoire of the octopus is broadly similar to that found across invertebrate bilaterians, except for massive expansions in two gene families previously thought to be uniquely enlarged in vertebrates: the protocadherins, which regulate neuronal development, and the C2H2 superfamily of zinc-finger transcription factors. Extensive messenger RNA editing generates transcript and protein diversity in genes involved in neural excitability, as previously described⁷, as well as in genes participating in a broad range of other cellular functions. We identified hundreds of cephalopod-specific genes, many of which showed elevated expression levels in such specialized structures as the skin, the suckers and the nervous system. Finally, we found evidence for large-scale genomic rearrangements that are closely associated with transposable element expansions. Our analysis suggests that substantial expansion of a handful of gene families, along with extensive remodelling of genome linkage and repetitive content, played a critical role in the evolution of cephalopod morphological innovations, including their large and complex nervous systems.

Soft-bodied cephalopods such as the octopus (Fig. 1a) show remarkable morphological departures from the basic molluscan body plan, including dexterous arms lined with hundreds of suckers that function as specialized tactile and chemosensory organs, and an elaborate chromatophore system under direct neural control that enables rapid changes in appearance^{1,8}. The octopus nervous system is vastly modified in size and organization relative to other molluscs, comprising a circumesophageal brain, paired optic lobes and axial nerve cords in each arm^{2,3}. Together these structures contain nearly half a billion neurons, more than six times the number in a mouse brain^{2,9}. Extant coleoid cephalopods show extraordinarily sophisticated behaviours including complex problem solving, task-dependent conditional discrimination, observational learning and spectacular displays of camouflage^{1,10} (Supplementary Videos 1 and 2).

To explore the genetic features of these highly specialized animals, we sequenced the *Octopus bimaculoides* genome by a whole-genome shotgun approach (Supplementary Note 1) and annotated it using extensive transcriptome sequence from 12 tissues (Methods and Supplementary Note 2). The genome assembly captures more than

97% of expressed protein-coding genes and 83% of the estimated 2.7 gigabase (Gb) genome size (Methods and Supplementary Notes 1–3). The unassembled fraction is dominated by high-copy repetitive sequences (Supplementary Note 1). Nearly 45% of the assembled genome is composed of repetitive elements, with two bursts of transposon activity occurring ~25-million and ~56-million years ago (Mya) (Supplementary Note 4).

We predicted 33,638 protein-coding genes (Methods and Supplementary Note 4) and found alternate splicing at 2,819 loci, but no locus showed an unusually high number of splice variants (Supplementary Note 4). A-to-G discrepancies between the assembled genome and transcriptome sequences provided evidence for extensive mRNA editing by adenosine deaminases acting on RNA (ADARs). Many candidate edits are enriched in neural tissues⁷ and are found in a range of gene families, including ‘housekeeping’ genes such as the tubulins, which suggests that RNA edits are more widespread than previously appreciated (Extended Data Fig. 1 and Supplementary Note 5).

Based primarily on chromosome number, several researchers proposed that whole-genome duplications were important in the evolution of the cephalopod body plan^{4–6}, paralleling the role ascribed to the independent whole-genome duplication events that occurred early in vertebrate evolution¹¹. Although this is an attractive framework for both gene family expansion and increased regulatory complexity across multiple genes, we found no evidence for it. The gene family expansions present in octopus are predominantly organized in clusters along the genome, rather than distributed in doubly conserved synteny as expected for a paleopolyploid^{12,13} (Supplementary Note 6.2). Although genes that regulate development are often retained in multiple copies after paleopolyploidy in other lineages, they are not generally expanded in octopus relative to limpet, oyster and other invertebrate bilaterians^{11,14} (Table 1 and Supplementary Notes 7.4 and 8).

Hox genes are commonly retained in multiple copies following whole-genome duplication¹⁵. In *O. bimaculoides*, however, we found only a single Hox complement, consistent with the single set of Hox transcripts identified in the bobtail squid *Euprymna scolopes* with PCR¹⁶. Remarkably, octopus Hox genes are not organized into clusters as in most other bilaterian genomes¹⁵, but are completely atomized (Extended Data Fig. 2 and Supplementary Note 9). Although we cannot rule out whole-genome duplication followed by considerable gene loss, the extent of loss needed to support this claim would far exceed that which has been observed in other paleopolyploid lineages, and it is more plausible that chromosome number in coleoids increased by chromosome fragmentation.

Mechanisms other than whole-genome duplications can drive genomic novelty, including expansion of existing gene families, evolution of novel genes, modification of gene regulatory networks, and reorganization of the genome through transposon activity. Within the *O. bimaculoides* genome, we found evidence for all of these

¹Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637, USA. ²Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 9040495, Japan.

³Centre for Organismal Studies, University of Heidelberg, 69117 Heidelberg, Germany. ⁴Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. ⁵Department of Neurobiology, University of Chicago, Chicago, Illinois 60637, USA. ⁶Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA.

*These authors contributed equally to this work.

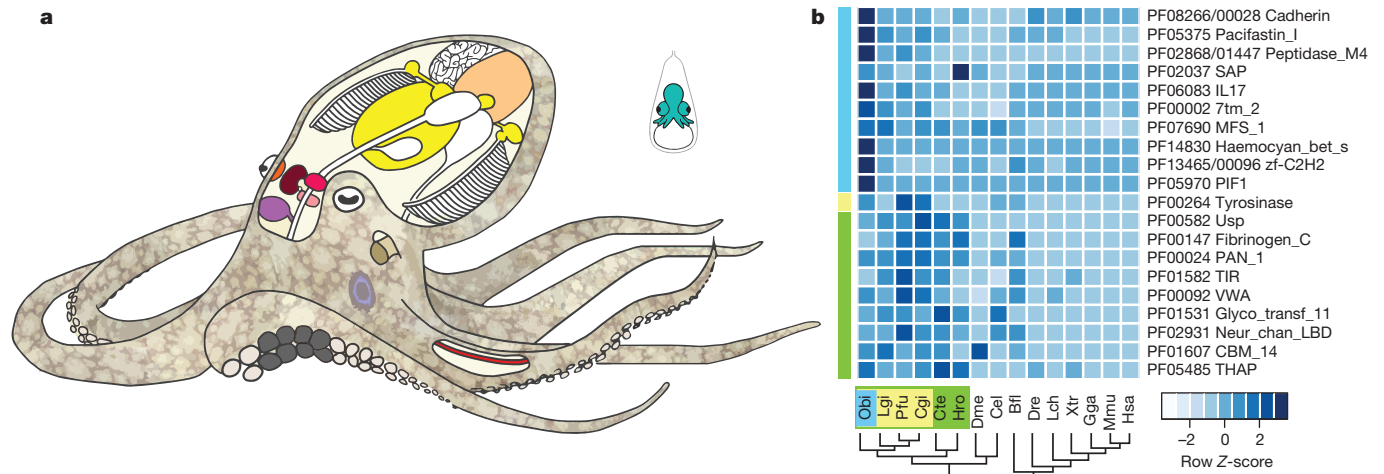


Figure 1 | Octopus anatomy and gene family representation analysis.

a, Schematic of *Octopus bimaculoides* anatomy, highlighting the tissues sampled for transcriptome analysis: viscera (heart, kidney and hepatopancreas), yellow; gonads (ova or testes), peach; retina, orange; optic lobe (OL), maroon; supraesophageal brain (Supra), bright pink; subesophageal brain (Sub), light pink; posterior salivary gland (PSG), purple; axial nerve cord (ANC), red; suckers, grey; skin, mottled brown; stage 15 (St15) embryo, aquamarine. Skin sampled for transcriptome analysis included the eyespot, shown in light blue. **b**, C2H2 and protocadherin domain-containing gene families are expanded in octopus. Enriched Pfam domains were identified in

lophotrochozoans (green) and molluscs (yellow), including *O. bimaculoides* (light blue). For a domain to be labelled as expanded in a group, at least 50% of its associated gene families need a corrected *P* value of 0.01 against the outgroup average. Some Pfams (for example, Cadherin and Cadherin_2) may occur in the same gene, however multiple domains in a given gene were counted only once. Bfl, *Branchiostoma floridae*; Cel, *Caenorhabditis elegans*; Cgi, *Crassostrea gigas*; Cte, *Capitella teleta*; Dme, *Drosophila melanogaster*; Dre, *Danio rerio*; Gga, *Gallus gallus*; Hsa, *Homo sapiens*; Hro, *Helobdella robusta*; Lch, *Latimeria chalumnae*; Lgi, *Lottia gigantea*; Mmu, *Mus musculus*; Obi, *O. bimaculoides*; Pfu, *Pinctada fucata*; Xtr, *Xenopus tropicalis*.

mechanisms, including expansions in several gene families, a suite of octopus- and cephalopod-specific genes, and extensive genome shuffling.

In gene family content, domain architecture and exon-intron structure, the octopus genome broadly resembles that of the limpet *Lottia gigantea*¹⁷, the polychaete annelid *Capitella teleta*¹⁷ and the cephalochordate *Branchiostoma floridae*¹⁴ (Supplementary Note 7 and Extended Data Fig. 3). Relative to these invertebrate bilaterians, we found a fairly standard set of developmentally important transcription factors and signalling pathway genes, suggesting that the evolution of the cephalopod body plan did not require extreme expansions of these 'toolkit' genes (Table 1 and Supplementary Note 8.2). However, statistical analysis of protein domain distributions across animal genomes did identify several notable gene family expansions in octopus, including protocadherins, C2H2 zinc-finger proteins (C2H2 ZNFs), interleukin-17-like genes (IL17-like), G-protein-coupled receptors (GPCRs), chitinases and sialins (Figs 1b, 2 and 3; Extended Data Figs 4–6 and Supplementary Notes 8 and 10).

The octopus genome encodes 168 multi-exonic protocadherin genes, nearly three-quarters of which are found in tandem clusters on the genome (Fig. 2b), a striking expansion relative to the 17–25 genes found in *Lottia*, *Crassostrea gigas* (oyster) and *Capitella* genomes. Protocadherins are homophilic cell adhesion molecules whose function has been primarily studied in mammals, where they are required for neuronal development and survival, as well as synaptic specificity¹⁸. Single protocadherin genes are found in the invertebrate deuterostomes *Saccoglossus kowalevskii* (acorn worm) and *Strongylocentrotus purpuratus* (sea urchin), indicating that their absence in *Drosophila melanogaster* and *Caenorhabditis elegans* is due to gene loss. Vertebrates also show a remarkable expansion of the protocadherin repertoire, which is generated by complex splicing from a clustered locus rather than tandem gene duplication (reviewed in ref. 19). Thus both octopuses and vertebrates have independently evolved a diverse array of protocadherin genes.

A search of available transcriptome data from the longfin inshore squid *Doryteuthis* (formerly, *Loligo*) *pealeii*²⁰ also demonstrated an expanded number of protocadherin genes (Supplementary Note 8.3). Surprisingly, our phylogenetic analyses suggest that the squid

and octopus protocadherin arrays arose independently. Unlinked octopus protocadherins appear to have expanded ~135 Mya, after octopuses diverged from squid. In contrast, clustered octopus protocadherins are much more similar in sequence, either due to more recent duplications or gene conversion as found in clustered protocadherins in zebrafish and mammals²¹.

The expression of protocadherins in octopus neural tissues (Fig. 2) is consistent with a central role for these genes in establishing and maintaining cephalopod nervous system organization as they do in vertebrates. Protocadherin diversity provides a mechanism for regulating the short-range interactions needed for the assembly of local neural circuits¹⁸, which is where the greatest complexity in the cephalopod nervous system appears². The importance of local neuropil interactions, rather than long-range connections, is probably due to the limits placed on axon density and connectivity by the absence of myelin, as thick axons are then required for rapid high-fidelity signal conduction over long distances. The sequence divergence between octopus and

Table 1 | Metazoan developmental control genes

	Obi	Lgi	Cte	Dme	Cel	Bfl	Hsa
Ligands							
Fibroblast growth factor	3	2	1	3	3	8	22
Wnt	12	10	12	7	5	17	19
TGFβ/BMP	12	9	14	6	5	22	33
Delta/Jagged	4	1	1	2	4	2	7
Hedgehog	1	1	1	1	0	1	3
Axon guidance	10	9	9	6	8	23	33
Transcription factors							
C2H2 zinc-finger	1,790	413	222	326	211	1,338	764
Homeodomain	114	121	111	104	99	133	333
High mobility group	23	15	14	13	16	51	125
Helix loop helix	50	63	64	59	42	78	118
Nuclear hormone receptor	40	44	45	16	274	33	48
Fox	16	28	26	17	18	42	43
Tbox	9	9	7	8	21	9	18

Number of members of developmental ligand and transcription factor families from *O. bimaculoides* and selected other taxa. Dendrogram above species names reflects their evolutionary relationships. Bfl, *Branchiostoma floridae*; Cel, *Caenorhabditis elegans*; Cte, *Capitella teleta*; Dme, *Drosophila melanogaster*; Hsa, *Homo sapiens*; Lgi, *Lottia gigantea*; Obi, *O. bimaculoides*.

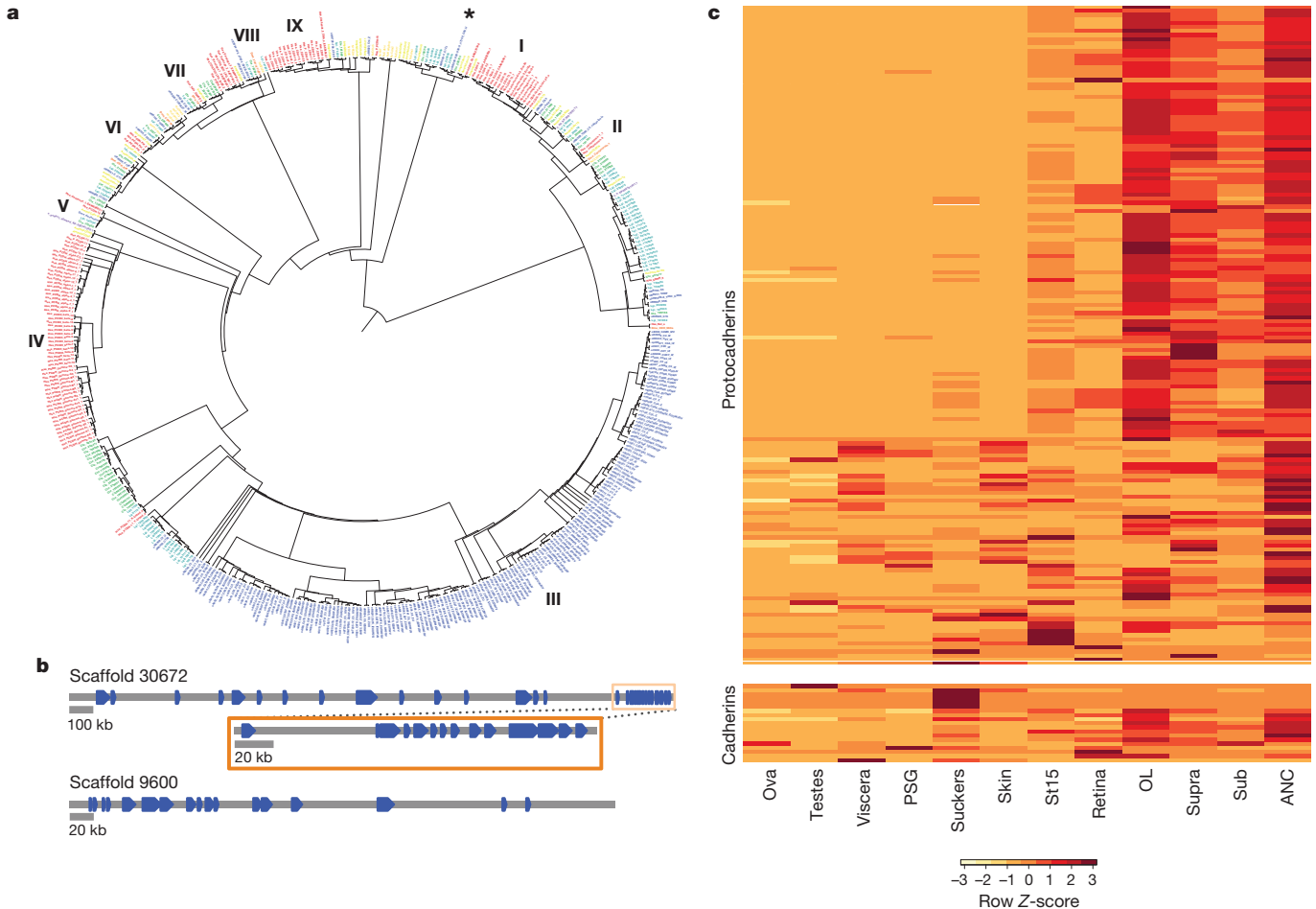


Figure 2 | Protocadherin expansion in octopus. **a**, For a larger version of panel **a**, see Extended Data Fig. 11. Phylogenetic tree of cadherin genes in Hsa (red), Dme (orange), *Nematostella vectensis* (mustard yellow), *Amphimedon queenslandica* (yellow), Cte (green), Lgi (teal), Obi (blue), and *Saccoglossus kowalevskii* (purple). I, Type I classical cadherins; II, calyntenins; III, octopus protocadherin expansion (168 genes); IV, human protocadherin expansion (58 genes); V, dachsous; VI, fat-like; VII, fat; VIII, CELSR; IX, Type II classical cadherins. Asterisk denotes a novel cadherin with over 80 extracellular cadherin domains found in Obi and Cte. **b**, Scaffold 30672 and Scaffold 9600

contain the two largest clusters of protocadherins, with 31 and 17, respectively. Clustered protocadherins vary greatly in genomic span and are oriented in a head-to-tail manner along each scaffold. **c**, Expression profiles of 161 protocadherins and 19 cadherins in 12 octopus tissues; 7 protocadherins were not detected in the tissues sampled. Cells are coloured according to number of standard deviations from the mean expression level. Protocadherins have high expression in neural tissues. Cadherins generally show a similar expression pattern, with the exception of a group of sucker-specific cadherins.

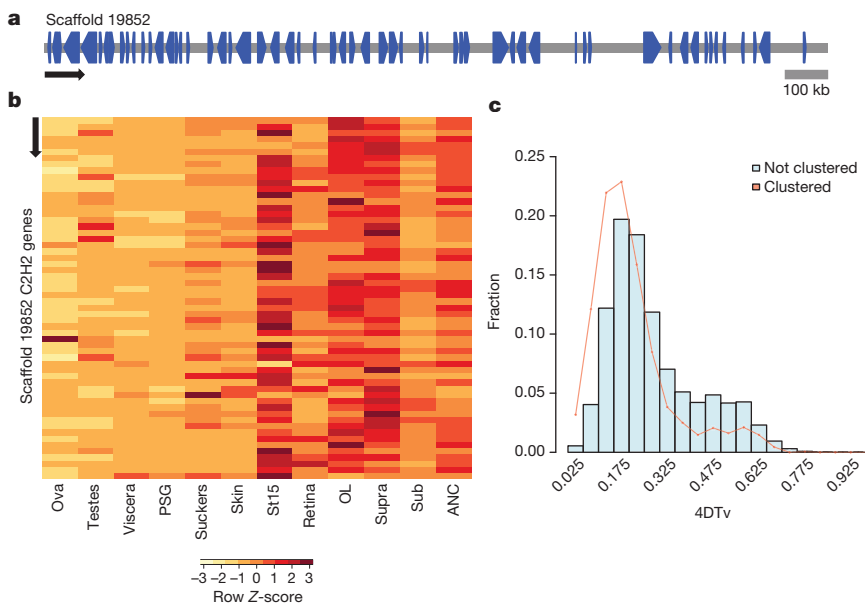


Figure 3 | C2H2 ZNF expansion in octopus.

a, Genomic organization of the largest C2H2 cluster. Scaffold 19852 contains 58 C2H2 genes that are transcribed in different directions. **b**, Expression profile of C2H2 genes along Scaffold 19852 in 12 octopus transcriptomes. Neural and developmental transcriptomes show high levels of expression for a majority of these C2H2 genes. In **a** and **b**, arrow denotes scaffold orientation. **c**, Distribution of fourfold synonymous site transversion distances (4DTv) between C2H2-domain-containing genes.

squid protocadherin expansions may reflect the notable differences between octopuses and decapodiforms in brain organization, which have been most clearly demonstrated for the vertical lobe, a key structure in cephalopod learning and memory circuits^{2,22}. Finally, the independent expansions and nervous system enrichment of protocadherins in coleoid cephalopods and vertebrates offers a striking example of convergent evolution between these clades at the molecular level.

As with the protocadherins, we found multiple clusters of C2H2 ZNF transcription factor genes (Fig. 3a and Supplementary Note 8.4). The octopus genome contains nearly 1,800 multi-exonic C2H2-containing genes (Table 1), more than the 200–400 C2H2 ZNFs found in other lophotrochozoans and the 500–700 found in eutherian mammals, in which they form the second-largest gene family²³. C2H2 ZNF transcription factors contain multiple C2H2 domains that, in combination, result in highly specific nucleic acid binding. The octopus C2H2 ZNFs typically contain 10–20 C2H2 domains but some have as many as 60 (Supplementary Note 8.4). The majority of the transcripts are expressed in embryonic and nervous tissues (Fig. 3b). This pattern of expression is consistent with roles for C2H2 ZNFs in cell fate determination, early development and transposon silencing, as demonstrated in genetic model systems²³.

The expansion of the *O. bimaculoides* C2H2 ZNFs coincides with a burst of transposable element activity at ~25 Mya (Fig. 3c). The flanking regions of these genes show a significant enrichment in a 70–90 base pair (bp) tandem repeat (31% for C2H2 genes versus 4% for all genes; Fisher's exact test P value $<1 \times 10^{-16}$), which parallels the linkage of C2H2 gene expansions to β -satellite repeats in humans²⁴. We also found an expanded C2H2 ZNF repertoire in amphioxus (Table 1), showing a similar enrichment in satellite-like repeats. These parallels suggest a common mode of expansion of a highly dynamic transcription factor family implicated in lineage-specific innovations.

To investigate further the evolution of gene families implicated in nervous system development and function, we surveyed genes associated with axon guidance (Table 1) and neurotransmission (Table 2), identifying their homologues in octopus and comparing numbers across a diverse set of animal genomes (Supplementary Notes 8–10). Several patterns emerged from this survey. The gene complements present in the model organisms *D. melanogaster* and *C. elegans* often showed striking departures from those seen in lophotrochozoans and vertebrates (Table 2 and Supplementary Note 10). For example, *D. melanogaster* encodes one member of the discs large (DLG) family, a key component of the postsynaptic scaffold. In contrast, mammals have four DLGs, which (along with other observations) led to suggestions that vertebrates possess uniquely complex synaptic machinery²⁵. However, we found three DLGs in both octopus and limpet, suggesting that vertebrate and fly gene number differences are not necessarily diagnostic of exceptional vertebrate synaptic complexity (Supplementary Note 10.6).

Overall, neurotransmission gene family sizes in the octopus were very similar to those seen in other lophotrochozoans (Table 2 and Supplementary Note 10), except for a few strikingly expanded gene families such as the sialic acid vesicular transporters (sialins) (Supplementary Note 10.2). We did find variations in the sizes of neurotransmission gene families between human and lophotrochozoans (Table 2 and Supplementary Note 10), but no evidence for systematic expansion of these gene families in vertebrates relative to octopus or other lophotrochozoans. Although some gene families were larger in mammals or absent in lophotrochozoans (for example, ligand-gated 5-HT receptors), others were absent in mammals and present in invertebrates (for example, anionic glutamate and acetylcholine receptors). The complement of neurotransmission genes in octopus may be broadly typical for a lophotrochozoan, but our findings suggest it is also not obviously smaller than is found in mammals.

Among the octopus complement of ligand-gated ion channels, we identified a set of atypical nicotinic acetylcholine receptor-like genes,

most of which are tandemly arrayed in clusters (Extended Data Fig. 7). These subunits lack several residues identified as necessary for the binding of acetylcholine²⁶, so it is unlikely that they function as acetylcholine receptors. The high level of expression of these divergent subunits within the suckers raises the interesting possibility that they act as sensory receptors, as do some divergent glutamate receptors in other protostomes²⁷. In addition, we identified 74 *Aplysia*-like and 11 vertebrate-like candidate chemoreceptors among the octopus GPCR superfamily of ~330 genes (Extended Data Fig. 6).

We found, amid extensive transcription of octopus transposons, that a class of octopus-specific short interspersed nuclear element sequences (SINEs) is highly expressed in neural tissues (Supplementary Note 4 and Extended Data Fig. 8). Although the role of active transposons is unclear, elevated transposon expression in neural tissues has been suggested to serve an important function in learning and memory in mammals and flies²⁸.

Transposable element insertions are often associated with genomic rearrangements²⁹ and we found that the transposon-rich octopus genome displays substantial loss of ancestral bilaterian linkages that are conserved in other species (Supplementary Note 6 and Extended Data Fig. 9). Interestingly, genes that are linked in other bilaterians but not in octopus are enriched in neighbouring SINE content. SINE insertions around these genes date to the time of tandem C2H2 expansion (Extended Data Fig. 9d), pointing to a crucial period of genome evolution in octopus. Other transposons such as Mariner show no such enrichment, suggesting distinct roles for different classes of transposons in shaping genome structure (Extended Data Fig. 9c).

Transposable element activity has been implicated in the modification of gene regulation across several eukaryotic lineages²⁹. We found that in the nervous system, the degree to which a gene's expression is tissue-specific is positively correlated with the transposon load around that gene (r^2 values ranging from 0.49 in the optic lobe to 0.81 in the subesophageal brain; Extended Data Fig. 8 and Supplementary Note 4). This correlation may reflect modulation of gene expression by transposon-derived enhancers or a greater tolerance for transposon insertion near genes with less complex patterns of tissue-specific gene regulation.

Using a relaxed molecular clock, we estimate that the octopus and squid lineages diverged ~270 Mya, emphasizing the deep evolutionary history of coleoid cephalopods^{8,30} (Supplementary Note 7.1 and Extended Data Fig. 10a). Our analyses found hundreds of coleoid- and octopus-specific genes, many of which were expressed in tissues containing novel structures, including the chromatophore-laden skin, the suckers and the nervous system (Extended Data Fig. 10 and Supplementary Note 11). Taken together, these novel genes, the

Table 2 | Ion channel subunits

	Obi	Aca	Lgi	Cte	Dme	Cel	Hsa
Voltage-gated calcium channels	8	8	6	10	9	10	10
Voltage-gated sodium channels	3	2	3	2	4	0	13
Transient receptor potential channels	36	45	40	43	13	23	29
K⁺ channels							
Voltage-gated	30	23	29	20	10	51	40
Calcium-activated, small/large conductance	12	8	9	6	3	6	8
Inward rectifying	3	4	5	6	4	3	16
Two pore	12	9	12	14	11	47	15
Non-voltage-gated	27	21	26	26	18	72	39
Cys-loop receptors							
Glutamate	21	15	47	36	30	15	18
Nicotinic acetylcholine	53	16	52	77	10	88	16
Inhibitory acetylcholine	3	2	5	2	0	4	0
5-HT3	0	0	0	0	0	1	5
GABA	6	5	4	9	3	7	19
Glutamate-gated chloride channels	7	5	8	5	1	6	0

Number of subunits of representative ion channel families in *O. bimaculoides* and across examined taxa. Dendrogram above species names shows their evolutionary relationships. Aca, *Aplysia californica*.

expansion of C2H2 ZNFs, genome rearrangements, and extensive transposable element activity yield a new landscape for both *trans*- and *cis*-regulatory elements in the octopus genome, resulting in changes in an otherwise 'typical' lophotrochozoan gene complement that contributed to the evolution of cephalopod neural complexity and morphological innovations.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 December 2014; accepted 16 June 2015.

- Hanlon, R. T. & Messenger, J. B. *Cephalopod Behaviour* (Cambridge Univ. Press, 1996).
- Young, J. Z. *The Anatomy of the Nervous System of Octopus vulgaris* (Clarendon Press, 1971).
- Wells, M. J. *Octopus: Physiology and Behaviour of an Advanced Invertebrate* (Chapman and Hall, 1978).
- Bonnaud, L., Ozouf-Costaz, C. & Boucher-Rodoni, R. A molecular and karyological approach to the taxonomy of Nautilus. *C. R. Biol.* **327**, 133–138 (2004).
- Hallinan, N. M. & Lindberg, D. R. Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. *Genome Biol. Evol.* **3**, 1150–1163 (2011).
- Yoshida, M. A. *et al.* Genome structure analysis of molluscs revealed whole genome duplication and lineage specific repeat variation. *Gene* **483**, 63–71 (2011).
- Rosenthal, J. J. & Seeburg, P. H. A-to-I RNA editing: effects on proteins key to neural excitability. *Neuron* **74**, 432–439 (2012).
- Kröger, B., Vinther, J. & Fuchs, D. Cephalopod origin and evolution. *Bioessays* **33**, 602–613 (2011).
- Herculano-Houzel, S., Mota, B. & Lent, R. Cellular scaling rules for rodent brains. *Proc. Natl Acad. Sci. USA* **103**, 12138–12143 (2006).
- Grasso, F. W. & Basil, J. A. The evolution of flexible behavioral repertoires in cephalopod molluscs. *Brain Behav. Evol.* **74**, 231–245 (2009).
- Holland, P. W., Garcia-Fernandez, J., Williams, N. A. & Sidow, A. Gene duplications and the origins of vertebrate development. *Development (Suppl.)*, 125–133 (1994).
- Dietrich, F. S. *et al.* The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**, 304–307 (2004).
- Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).
- Putnam, N. H. *et al.* The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071 (2008).
- Duboule, D. The rise and fall of Hox gene clusters. *Development* **134**, 2549–2560 (2007).
- Callaerts, P. *et al.* HOX genes in the sepiolid squid *Euprymna scolopes*: implications for the evolution of complex body plans. *Proc. Natl Acad. Sci. USA* **99**, 2088–2093 (2002).
- Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526–531 (2013).
- Zipursky, S. L. & Sanes, J. R. Chemoaffinity revisited: Dscams, protocadherins, and neural circuit assembly. *Cell* **143**, 343–353 (2010).
- Chen, W. V. & Maniatis, T. Clustered protocadherins. *Development* **140**, 3297–3302 (2013).
- Brown, C. T., Graveley, B. & Rosenthal, J. J. *Loligo pealeii (Squid) Data Dump* (<http://ivory.idyll.org/blog/2014-loligo-transcriptome-data.html>) (2014).
- Noonan, J. P., Grimwood, J., Schmutz, J., Dickson, M. & Myers, R. M. Gene conversion and the evolution of protocadherin gene cluster diversity. *Genome Res.* **14**, 354–366 (2004).
- Shomrat, T. *et al.* Alternative sites of synaptic plasticity in two homologous “fan-out fan-in” learning and memory networks. *Curr. Biol.* **21**, 1773–1782 (2011).
- Liu, H., Chang, L. H., Sun, Y., Lu, X. & Stubbs, L. Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. *Genome Biol. Evol.* **6**, 510–525 (2014).
- Eichler, E. E. *et al.* Complex β -satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12. *Genome Res.* **8**, 791–808 (1998).
- Nithianantharajah, J. *et al.* Synaptic scaffold evolution generated components of vertebrate cognitive complexity. *Nature Neurosci.* **16**, 16–24 (2013).
- Brejci, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).
- Croset, V. *et al.* Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet.* **6**, e1001064 (2010).
- Erwin, J. A., Marchetto, M. C. & Gage, F. H. Mobile DNA elements in the generation of diversity and complexity in the brain. *Nature Rev. Neurosci.* **15**, 497–506 (2014).
- Chénais, B., Caruso, A., Hiard, S. & Casse, N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene* **509**, 7–15 (2012).
- Strugnell, J., Norman, M., Jackson, J., Drummond, A. J. & Cooper, A. Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach; the effect of data partitioning on resolving phylogenies in a Bayesian framework. *Mol. Phylogenet. Evol.* **37**, 426–441 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. T. Brown and J. Rosenthal for making *Doryteuthis* RNA-seq data available before publication; C. Ha, J. Orenstein, J. Brandenburger, M. Glotzer and H. Gui for bioinformatic assistance; S. Shigeno for help with tissue dissection; C. Huffard and R. Caldwell for providing the *O. bimaculoides* specimen used for genomic DNA isolation; and E. Begovic for genomic DNA preparation. This work was supported by the Molecular Genetics Unit of the Okinawa Institute of Science and Technology Graduate University (S.B. and D.S.R.) and by funding from the NSF (IOS-1354898) and NIH (R03 HD064887) to C.W.R. and from the NSF (DGE-0903637) to Z.Y.W. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 instrumentation grants S10RR029668 and S10RR027303, and the University of Chicago Functional Genomics Facility, supported by NIH grant UL1 TR000430.

Author Contributions The Chicago and the OIST/Berkeley groups initiated their transcriptome and genome projects independently. In the subsequent collaboration, both groups worked closely on every aspect of the project. Chicago group: C.B.A., Z.Y.W., J.R.P. and C.W.R.; OIST/Berkeley group: O.S., T.M., E.E.-G., S.B. and D.S.R.

Author Information Genome and transcriptome sequence reads have been deposited in the SRA as BioProjects PRJNA270931 and PRJNA285380. A browser of this genome assembly is available at (<http://octopus.metazome.net/>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.W.R. (cragsdale@uchicago.edu) or D.S.R. (dsrokhars@gmail.com).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

METHODS

Data access. Genome and transcriptome sequence reads are deposited in the SRA as BioProjects PRJNA270931 and PRJNA285380. The genome assembly and annotation are linked to the same BioProject ID. A browser of this genome assembly is available at (<http://octopus.metazome.net/>).

Genome sequencing and assembly. Genomic DNA from a single male *Octopus bimaculoides*³¹ was isolated and sequenced using Illumina technology to 60-fold redundant coverage in libraries spanning a range of pairs from ~350 bp to 10 kilobases (kb). These data were assembled with meraculous³² achieving a contig N50-length of 5.4 kb and a scaffold N50-length of 470 kb. The longest scaffold contains 99 genes and half of all predicted genes are on scaffolds with 8 or more genes (Supplementary Note 1).

Genome size and heterozygosity. The *O. bimaculoides* haploid genome size was estimated to be ~2.7 gigabases (Gb) based on fluorescence (2.66–2.68 Gb) and *k*-mer (2.86 Gb) measurements (Supplementary Notes 1 and 2), making it several times larger than other sequenced molluscan and lophotrochozoan genomes¹⁷. We observed nucleotide-level heterozygosity within the sequenced genome to be 0.08%, which may reflect a small effective population size relative to broadcast-spawning marine invertebrates.

Transcriptome sequencing. Twelve transcriptomes were sequenced from RNA isolated from ova, testes, viscera, posterior salivary gland (PSG), suckers, skin, developmental stage 15 (St15)³³, retina, optic lobe (OL), supraesophageal brain (Supra), subesophageal brain (Sub), and axial nerve cord (ANC) (Supplementary Note 2). RNA was isolated using TRIzol (Invitrogen) and 100-bp paired-end reads (insert size 300 bp) were generated on an Illumina HiSeq2000 sequencing machine.

De novo transcriptome assembly. Adapters and low-quality reads were removed before assembling transcriptomes using the Trinity *de novo* assembly package (version r2013-02-25 (refs 34, 35)). Assembly statistics are summarized in Supplementary Table 2.2. Following assembly, peptide-coding regions were translated using TransDecoder in the Trinity package. We compared the *de novo* assembled RNA-seq output to the genome to evaluate the completeness of the genome assembly. To minimize the number of spuriously assembled transcripts, only transcripts with ORFs predicted by TransDecoder were mapped onto the genome with BLASTN. Only 1,130 out of 48,259 transcripts with ORFs (2.34%) did not have a match in the genome with a minimum identity of 95%.

Annotation of transposable elements. Transposable elements were identified with RepeatScout and RepeatModeler³⁶, and the masking was done with RepeatMasker³⁷, as outlined in Supplementary Note 4.2. The most abundant transposable element is a previously identified octopus-specific SINE³⁸ that accounts for 4% of the assembled genome.

Annotation of protein-coding genes. Protein-coding genes were annotated by combining transcriptome evidence with homology-based and *de novo* gene prediction methods (Supplementary Note 4). For homology prediction we used predicted peptide sets of three previously sequenced molluscs (*L. gigantea*, *C. gigas*, and *A. californica*) along with selected other metazoans. Alternative splice isoforms were identified with PASA³⁹. Annotation statistics are provided in Supplementary Table 4.1.1. Genes known in vertebrates to have many isoforms, such as ankyrin, *TRAK1* and *LRCH1*, also show alternative splicing in octopus but at a more limited level. Octopus genes with ten or more alternative splice forms are provided in Supplementary Table 4.1.2.

Calibration of sequence divergence with respect to time. The divergence between squid and octopus was estimated using r8s⁴⁰ by fixing cephalopod divergence from bivalves and gastropods to 540 Mya⁸. Our estimate of 270 Mya for the squid–octopus divergence corresponds to mean neutral substitution rate of dS ~2 based on the protein-directed CDS alignments between the species (Supplementary Fig. 6.1.2) and a dS estimation using the yn00 program⁴¹. Throughout the manuscript we convert from sequence divergence to time by assuming that dS ~1 corresponds to 135 million years. For example, unlinked octopus protocadherins appear to have expanded ~135 Mya based on mean pairwise dS ~1, after octopuses diverged from squid. In contrast, clustered octopus protocadherins are much more similar in sequence (mean pairwise dS ~0.4, or ~55 Mya).

Quantifying gene expression. Transcriptome reads were mapped to the genome assembly with TopHat 2.0.11 (ref. 42). A range of 76–90% of reads from the different samples mapped to the genome. Mapped reads were sorted and indexed with SAMtools⁴³. The read counts in each tissue were produced with BEDTools multicov program⁴⁴ using the gene model coordinates. The counts were normalized by the total transcriptome size of each tissue and by the length of the gene. Heat maps showing expression patterns were generated in R using the heatmap.2 function.

Gene complement. Gene families of particular interest, including developmental regulatory genes, neural-related genes, and gene families that appear to be

expanded in *O. bimaculoides*, were manually curated and analysed. We searched the octopus genome and transcriptome assemblies using BLASTP and TBLASTN with annotated sequences from human, mouse and *D. melanogaster*. Bulk analyses were also performed using Pfam⁴⁵ and PANTHER⁴⁶. We used BLASTP and TBLASTX to search for specific gene families in deposited genome and transcriptome databases for *L. gigantea*, *A. californica*, *C. gigas*, *C. teleta*, *T. castaneum*, *D. melanogaster*, *C. elegans*, *N. vectensis*, *A. queenslandica*, *S. kowalevskii*, *B. floridae*, *C. intestinalis*, *D. rerio*, *M. musculus* and *H. sapiens*. Candidate genes were verified with BLAST⁴⁷ and Pfam⁴⁵ analysis. Genes identified in the octopus genome were confirmed and extended using the transcriptomes. Multiple gene models that matched the same transcript were combined. The identified sequences from octopus and other bilaterians were aligned with either MUSCLE⁴⁸, CLUSTALO⁴⁹, MacVector 12.6 (MacVector, North Carolina), or Jalview⁵⁰. Phylogenetic trees were constructed with FastTree⁵¹ using the Jones–Taylor–Thornton model of amino acid evolution, and visualized with FigTree v1.3.1.

Synten. Microsynteny was computed based on metazoan node gene families (Supplementary Note 7). We used Nmax 10 (maximum of 10 intervening genes) and Nmin 3 (minimum of three genes in a syntenic block) according to the pipeline described in ref. 17 (Supplementary Note 6). To simplify gene family assignments we limited our analyses to 4,033 gene families shared among human, amphioxus, *Capitella*, *Helobdella*, *Octopus*, *Lottia*, *Crassostrea*, *Drosophila* and *Nematostella*. We required ancestral bilaterian syntenic blocks to have a minimum of one species present in both ingroups, or in one ingroup and one outgroup. To examine the effect of fragmented genome assemblies, we simulated shorter assemblies by artificially fragmenting genomes to contain on average 5 genes per scaffold (Supplementary Note 6).

In comparison with other bilaterian genomes, we find that the octopus genome is substantially rearranged. In looking at microsyntenic linkages of genes with a maximum of 10 intervening genes, we found that octopus conserves only 34 out of 198 ancestral bilaterian microsyntenic blocks; the limpet *Lottia* and amphioxus retain more than twice as many such linkages (96 and 140, respectively). This difference remains significant after accounting for genes missed through orthology assignment as well as simulations of shorter scaffold sizes (Supplementary Note 6; Extended Data Fig. 9b). Scans for intra-genomic synteny, and doubly conserved synteny with *Lottia*, were performed as described in Supplementary Note 6.

Transposable elements and synten dynamics. The 5 kb upstream and downstream regions of genes were surveyed for transposable element (TE) content. For genes with non-zero TE load, their assignment to either conserved or lost bilaterian synteny in octopus was done using the microsynteny calculation described above. The number of genes for each category and TE class were as follows: 484 genes for retained synteny and 1,193 genes in lost synteny for all TE classes; 440 and 1,107, respectively, for SINEs; and 116 and 290, respectively, for Mariner. Wilcoxon *U*-tests for the difference of TE load in linked versus non-linked genes were conducted in R.

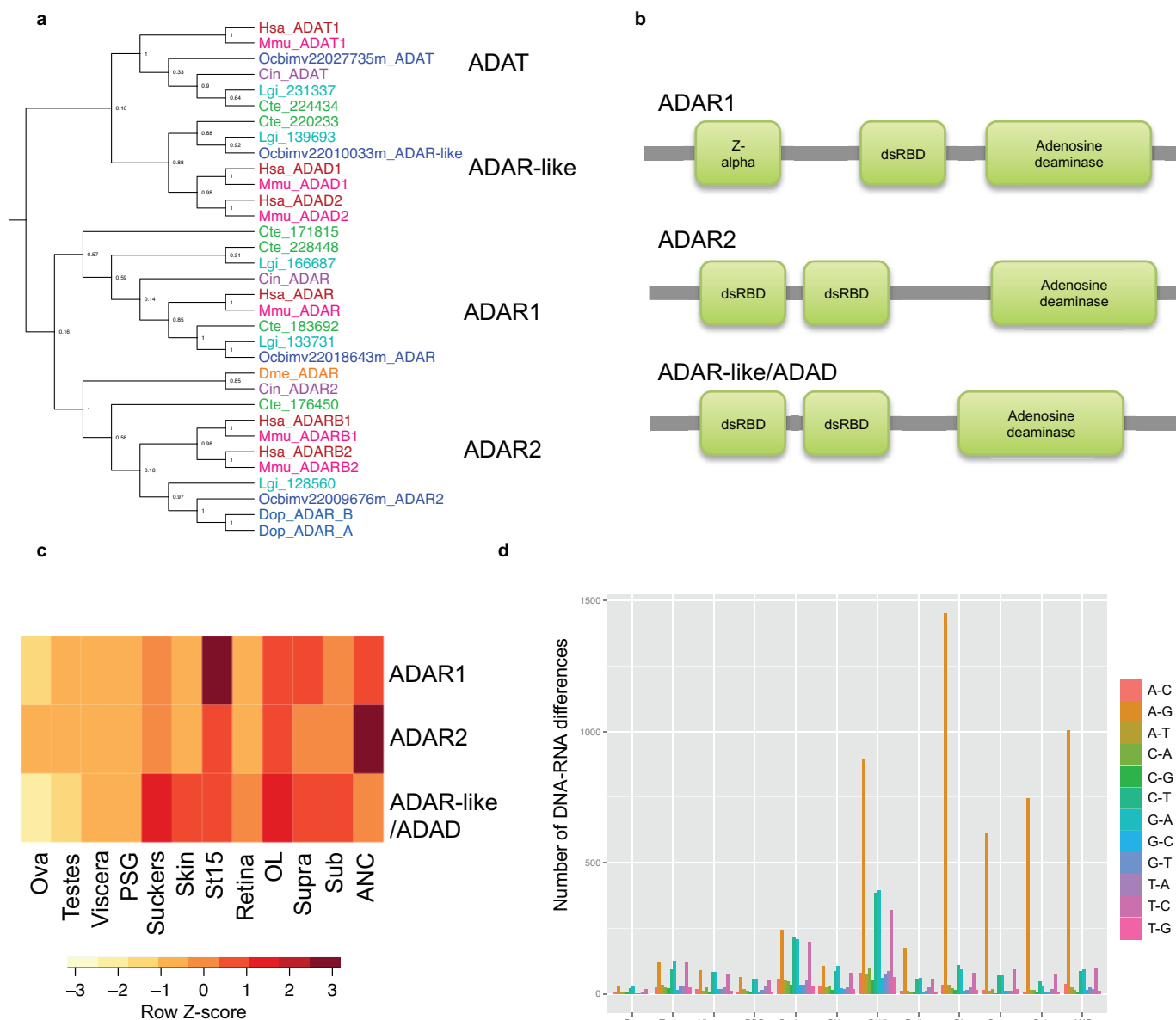
To assess transposon activity we assigned transcriptome reads aligned to 5,496,558 annotated transposon loci using BEDTools⁴⁴. Of these, 2,685,265 loci showed expression in at least one of the tissues.

RNA editing. RNA-seq reads were mapped to the genome with TopHat⁵², and SAMtools⁴³ was used to identify SNPs between the genomic and the RNA sequences. To identify polymorphic positions in the genome, SNPs and indels were predicted using GATK HaplotypeCaller version 3.1-1 in discovery mode with a minimum Phred scaled probability score of 30, based on an alignment of the 350 bp and 500 bp genomic fragment libraries using BWA-MEM version 0.7.6a. Using BEDTools⁴⁴, we removed SNPs predicted in both the transcriptome and the genome and discarded SNPs that had a Phred score below 40 or were outside of predicted genes. SNPs were binned according to the type of nucleotide change and the direction of transcription. Candidate edited genes were taken as those having SNPs with A-to-G substitutions in the predicted mRNA transcripts.

Cephalopod-specific genes. Cephalopod novelties were obtained by BLASTP and TBLASTN searches against the whole NR database⁵³ and a custom database of several mollusc transcriptomes (Supplementary Note 11.1). To ensure that we had as close to full-length sequence as possible, we extended proteins predicted from octopus genomic sequence with our *de novo* assembled transcriptomes, using the longest match to query NR, transcriptome and EST sequences from other animals. Gene sequences with transcriptome support but without a match to non-cephalopod animals at an e-value cutoff of 1×10^{-3} were considered for further analysis. Octopus sequences with a match of 1×10^{-5} or better to a sequence from another cephalopod were used to construct gene families, which were characterized by their BLAST alignments, HMM, PFAM-A/B, and UNIREF90 hits. The cephalopod-specific gene families are listed in the Source

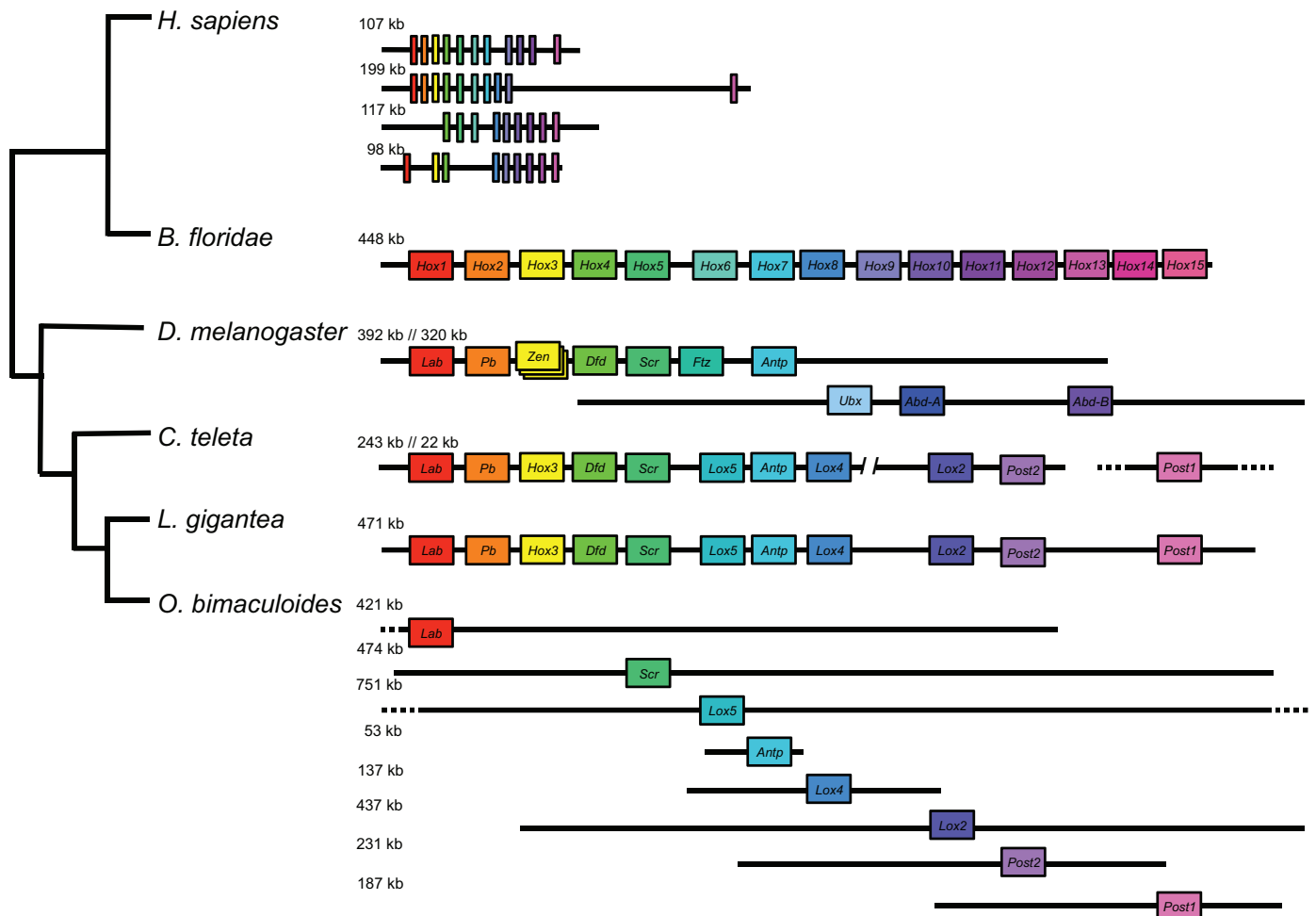
Data file for Extended Data Fig. 10. Octopus-specific novelties were defined as sequences with transcriptome support but without any matches to sequences from any other animals ($<1 \times 10^{-3}$), including nautiloid and decapodiform cephalopods.

31. Pickford, G. E. & McConnaughey, B. H. The *Octopus bimaculatus* problem: a study in sibling species. *Bulletin of the Bingham Oceanographic Collection* **12**, 1–66 (1949).
32. Chapman, J. A. *et al.* Meraculous: *de novo* genome assembly with short paired-end reads. *PLoS ONE* **6**, e23501 (2011).
33. Naef, A., Boletzky, S. v. & Roper, C. F. E. *Cephalopoda. Embryology* (Smithsonian Institution Libraries, 2000).
34. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnol.* **29**, 644–652 (2011).
35. Haas, B. J. *et al.* *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**, 1494–1512 (2013).
36. Smit, A. & Hubley, R. RepeatModeler Open-1.0. (2008–2010).
37. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-3.0. (1996–2010).
38. Ohshima, K. & Okada, N. Generality of the tRNA origin of short interspersed repetitive elements (SINEs). Characterization of three different tRNA-derived retrotransposons in the octopus. *J. Mol. Biol.* **243**, 25–37 (1994).
39. Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
40. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
41. Yang, Z. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43 (2000).
42. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
43. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
44. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
45. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
46. Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* **41**, D377–D386 (2013).
47. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
48. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
49. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
50. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
51. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
52. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**, 562–578 (2012).
53. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
54. Palavicini, J. P., O'Connell, M. A. & Rosenthal, J. J. An extra double-stranded RNA binding domain confers high activity to a squid RNA editing enzyme. *RNA* **15**, 1208–1218 (2009).
55. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
56. Starnes, T., Broxmeyer, H. E., Robertson, M. J. & Hromas, R. Cutting edge: IL-17D, a novel member of the IL-17 family, stimulates cytokine production and inhibits hemopoiesis. *J. Immunol.* **169**, 642–646 (2002).
57. Cummins, S. F. *et al.* Candidate chemoreceptor subfamilies differentially expressed in the chemosensory organs of the mollusc *Aplysia*. *BMC Biol.* **7**, 28 (2009).
58. van Nierop, P. *et al.* Identification of molluscan nicotinic acetylcholine receptor (nAChR) subunits involved in formation of cation- and anion-selective nAChRs. *J. Neurosci.* **25**, 10617–10626 (2005).



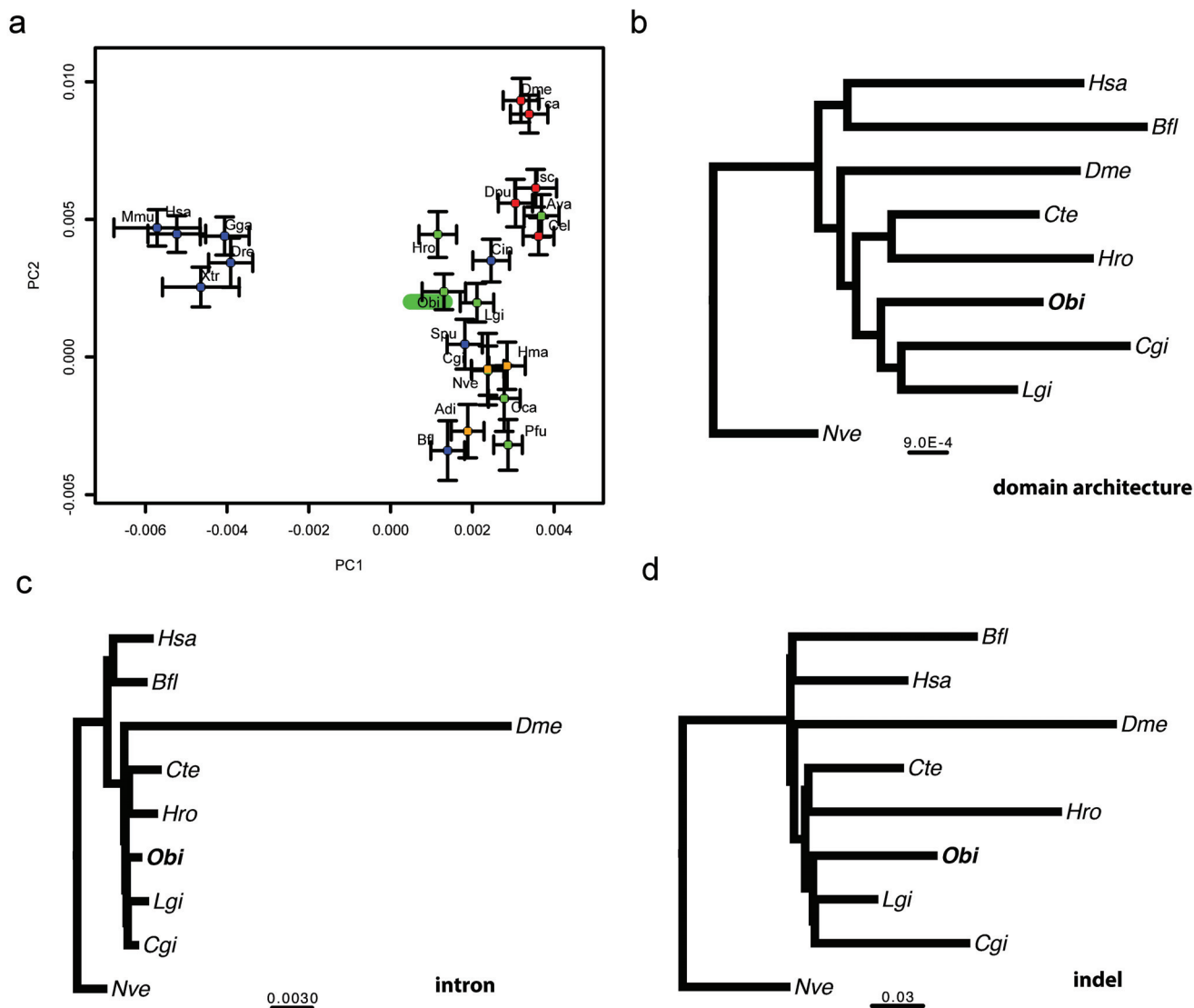
Extended Data Figure 1 | RNA editing in octopus. **a**, Approximate maximum likelihood tree of adenosine deaminases acting on RNA (ADARs) in bilaterians. ADAR1, ADAR2, ADAR-like/ADAD, and ADAT (tRNA-specific adenosine deaminase) were identified in Hsa, Mmu, Cin, Dme, Cte, Lgi, *D. opalescens* (Dop⁵⁴), and Obi with Shimodaira–Hasegawa-like support indicated at the nodes. **b**, *O. bimaculoides* ADAR1, ADAR2 and ADAR-like proteins contain one or two double-stranded RNA binding domains (dsRBD) as well as an adenosine deaminase domain. ADAR1 also has a z-alpha domain. **c**, Expression profiles of the three ADAR genes found in 12 *O. bimaculoides* tissues by RNA-seq profiling. **d**, DNA–RNA differences in *O.*

bimaculoides show prominent A-to-G changes. Histogram illustrates the number of DNA–RNA differences detected between coding sequences in the genome and 12 *O. bimaculoides* transcriptomes after filtering out polymorphisms identified in genomic sequencing. Differences were binned by the type of change (see key) in the direction of transcription. A-to-G changes are the most prevalent, particularly in neural tissues and during development, paralleling the expression of octopus ADARs in **c**. Other types of changes were also detected at lower levels, possibly resulting from uncharacterized polymorphisms.



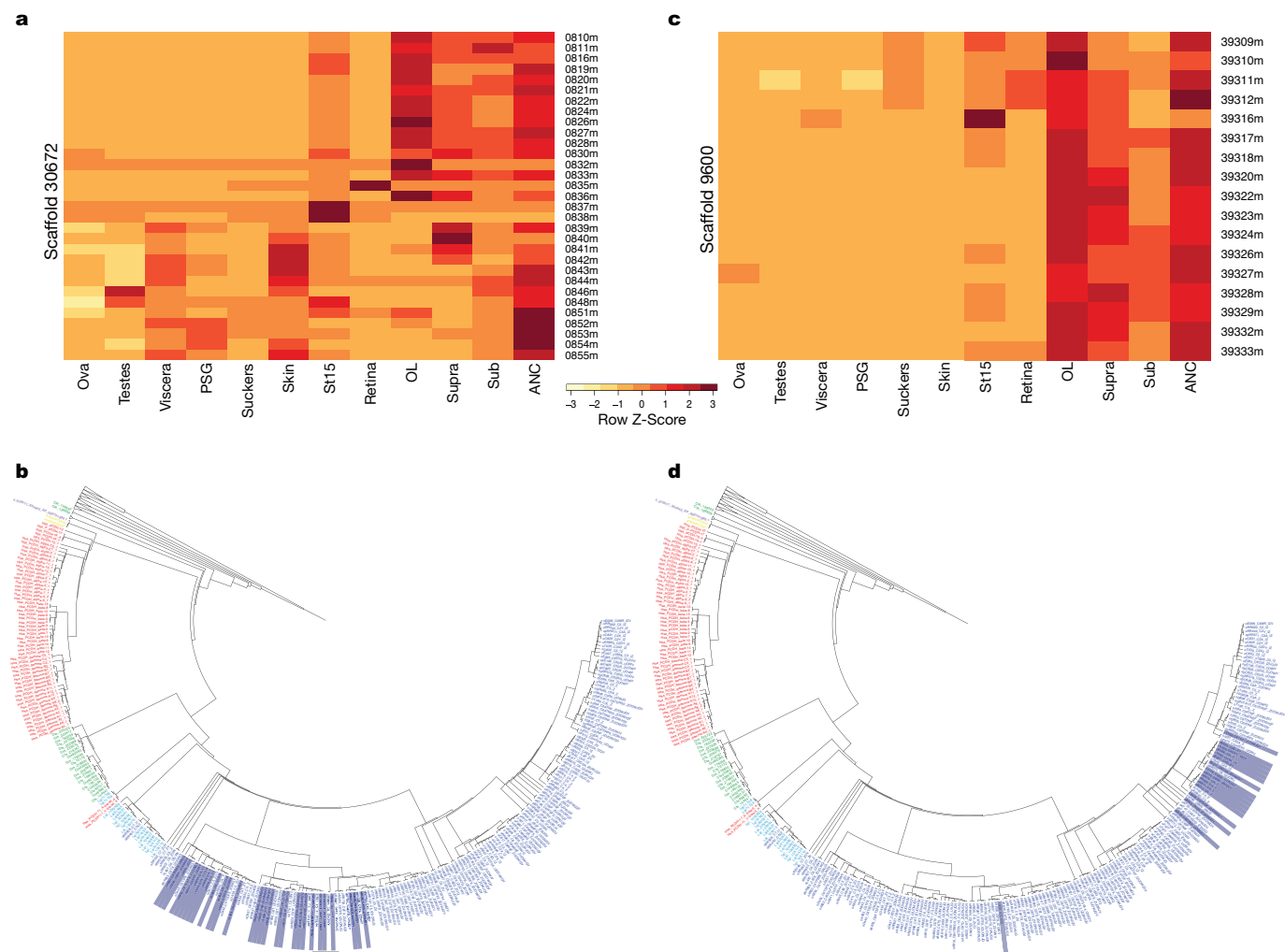
Extended Data Figure 2 | Local arrangement of Hox gene complement in *O. bimaculoides* and selected bilaterians. At the top, the four compact Hox clusters of *H. sapiens* and the single *B. floridae* cluster are depicted. The *D. melanogaster* Hox complex is split into two clusters. We included genes in the *D. melanogaster* locus that are homologues of Hox genes but have lost their homeotic function, such as *fushi tarazu* (*ftz*), *bicoid*, *zen* and *zen2* (the latter three are represented as overlapping boxes). Hox genes in *C. teleta* are found

on three scaffolds¹⁷. *L. gigantea* has a single cluster with the full known lophotrochozoan gene complement. In *O. bimaculoides* many of the scaffolds are several hundred kb long, and no two Hox genes are on the same scaffold. The positions of *O. bimaculoides* genes approximate their locations on scaffolds. Dashed lines indicate that the scaffold continues beyond what is shown. Scaffold length is depicted to scale with size noted on the left. Genes are positioned to illustrate orthology, which is also highlighted by colour.



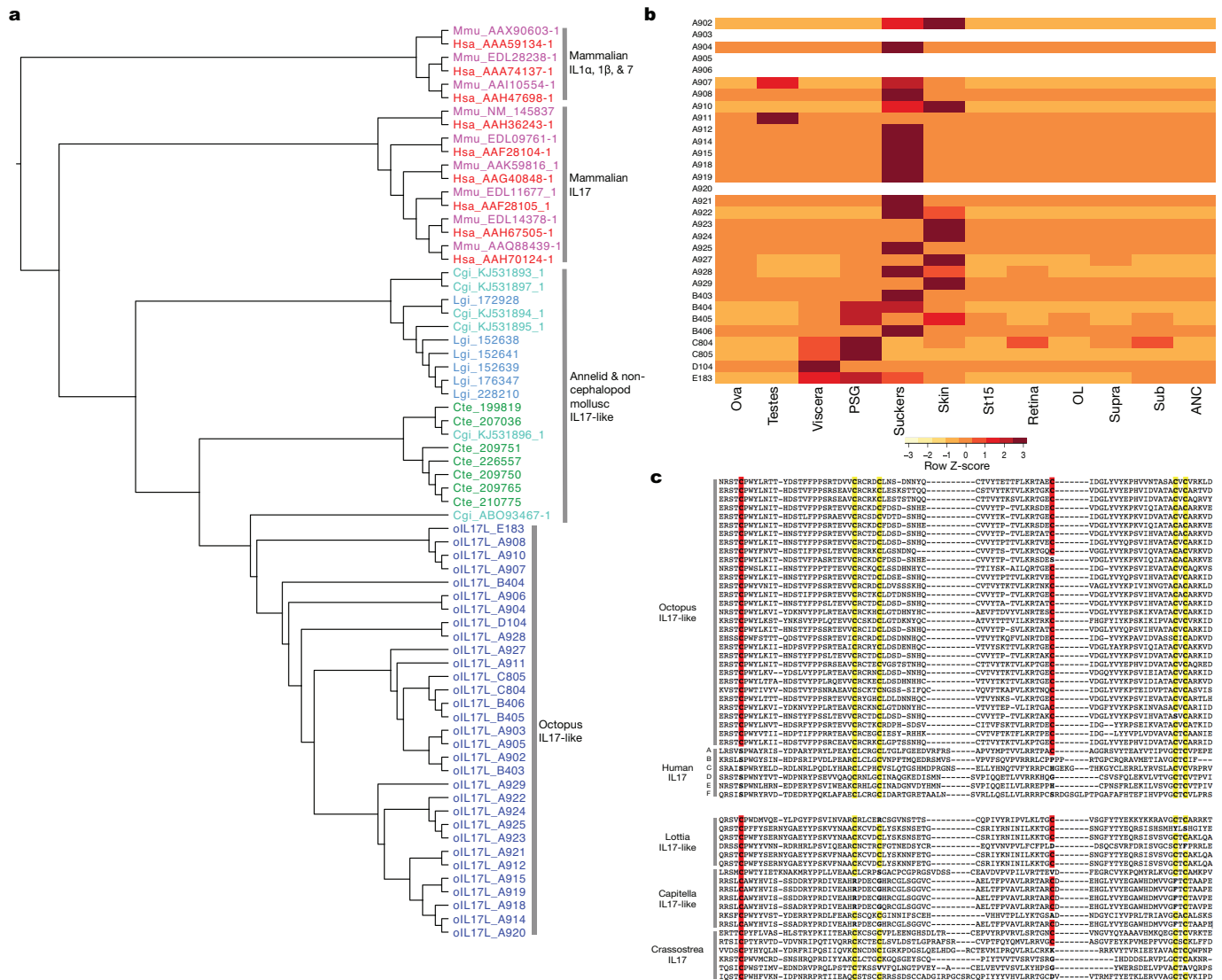
Extended Data Figure 3 | Gene complement and gene architecture evolution in metazoans. a, Principal component analysis of gene family counts. *O. bimaculoides* highlighted in green. Deuterostomes are indicated in blue, ecdysozoans in red, lophotrochozoans in green, and sponges and cnidarians in orange. Xtr, *Xenopus tropicalis*; Gga, *Gallus gallus*; Tca, *Tribolium castaneum*; Dpu, *Daphnia pulex*; Isc, *Ixodes scapularis*; Ava, *Adineta vaga*;

Spu, *S. purpuratus*; Hma, *Hydra magnipapillata*; Adi, *Acropora digitifera*. For methods, see Supplementary Note 7.4. **b-d**, MrBayes⁵⁵ tree (constrained topology) on binary characters of presence or absence of Pfam domain architectures (**b**), introns (**c**), or indels (**d**); scale bar represents estimated changes per site. For methods, see Supplementary Note 7.3.



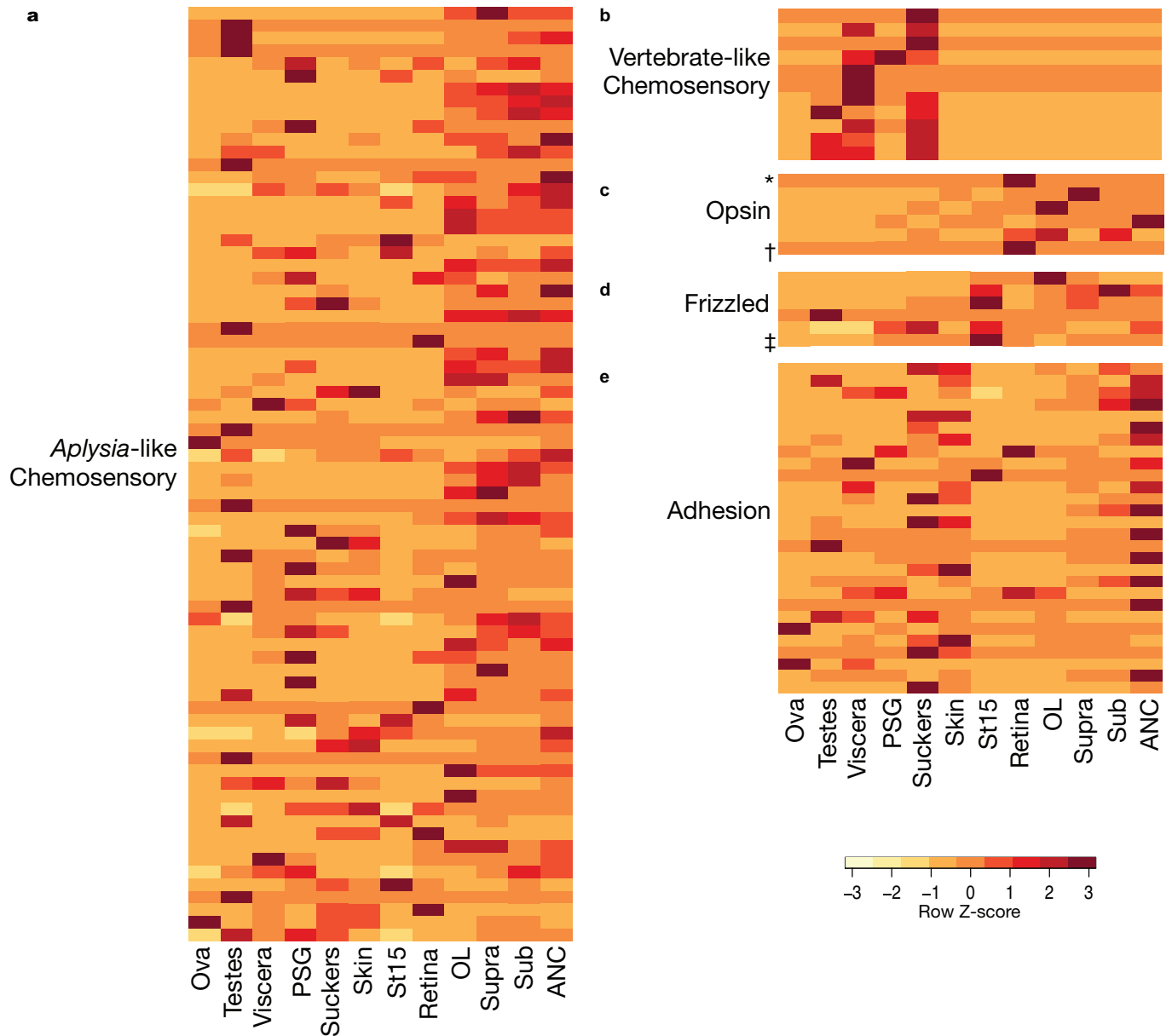
Extended Data Figure 4 | Protocadherin genes within a genomic cluster are similar in sequence and sites of expression. **a**, Expression profile of the 31 protocadherin genes located on Scaffold 30672 in 12 octopus transcriptomes. Over three-quarters of the protocadherins are highly expressed throughout central brain, OL and ANC, while the others show more mixed distributions. **b**, Phylogenetic tree highlighting Scaffold 30672 protocadherins in grey bars. **c**, Expression profile of the 17 protocadherin genes located on

Scaffold 9600. Almost all of these protocadherins are most highly expressed in nervous tissues, with the exception of Ocbimv220039316m, which is most highly expressed in the St15 sample. **d**, Phylogenetic tree highlighting Scaffold 9600 protocadherins in grey bars. As seen in **b**, protocadherins of the same scaffold tend to cluster together on the tree. Order of the genes in the heat maps (**a**, **c**) follows the ordering on the corresponding scaffold.



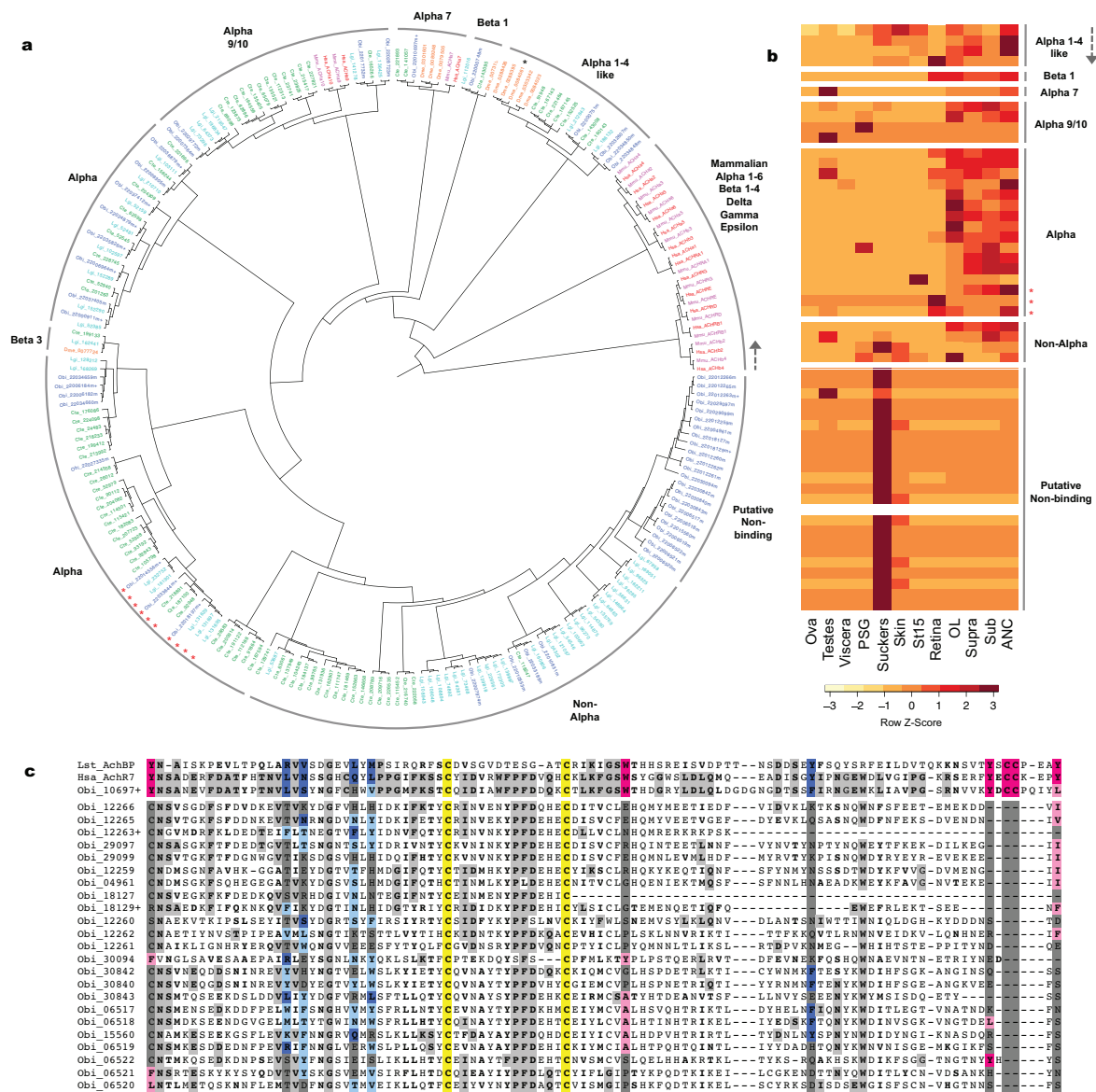
Extended Data Figure 5 | Expansion of interleukin 17 (IL17)-like genes.
a, Phylogenetic tree of interleukin genes in Obi, Cte, Cgi, and Lgi. Mammalian *IL1A*, *IL1B*, and *IL7* used as outgroups. Human and mouse *IL17*s branch from other members of the *IL* family. Octopus *IL*s (as well as all identified invertebrate *IL*s) group with the mammalian *IL17* branch and are named 'IL17-like'. The 31 octopus genes are distributed across 5 scaffolds: scaffold A (Obi_A), 23 members; scaffold B (Obi_B), 4 members; scaffold C (Obi_C), 2 members; scaffolds D (Obi_D) and E (Obi_E), 1 member each. **b**, Expression profile of 31 octopus IL17-like genes. Heat map rows are arranged by order on each scaffold. Blank rows indicate genes not expressed in our transcriptomes. The 27 genes found in our transcriptomes have strong expression in the suckers and skin. The scaffold C genes are enriched in the PSG

and the Scaffold D gene is enriched in the viscera. **c**, Conserved cysteine residues in human IL17 and invertebrate IL17-like proteins. The human IL17 proteins share a conserved cysteine motif comprising 4 cysteine residues, which may form interchain disulfide bonds and facilitate dimerization⁵⁶. Octopus IL17-like proteins also contain this four-cysteine motif, highlighted in yellow. One octopus sequence encodes only 3 of these highly conserved cysteine residues. These four cysteines are also present to varying degrees in *Lottia*, *Capitella* and *Crassostrea* sequences. Two additional conserved cysteine residues were found in the octopus sequences and are highlighted in red. The first cysteine residue is found in all invertebrate sequences examined, and none of the mammalian IL17 sequences.



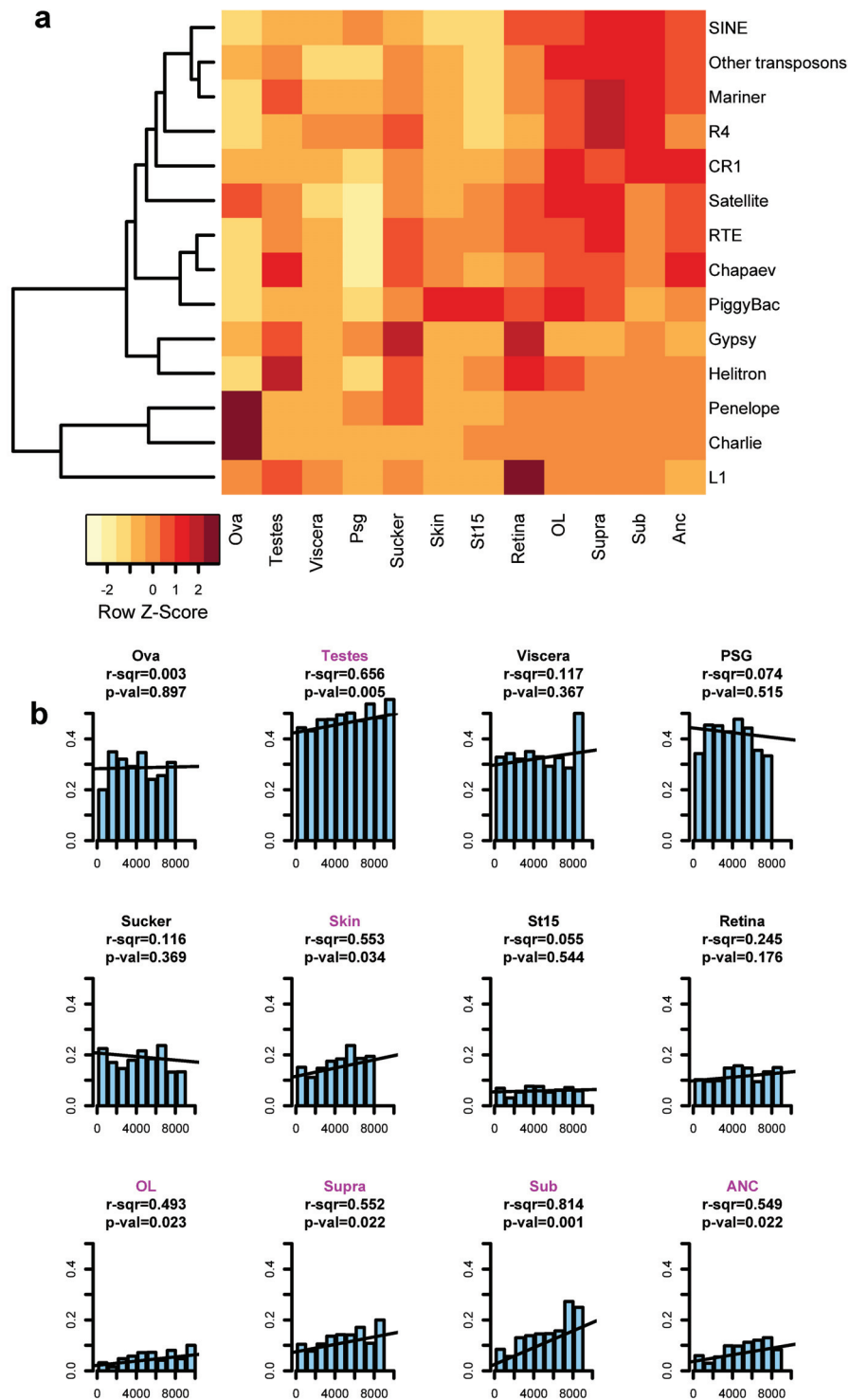
Extended Data Figure 6 | G-protein-coupled receptors. GPCRs, also known as 7-transmembrane (7TM) or serpentine receptors, form a large superfamily that activates intracellular second messenger systems upon ligand binding. This figure considers a subset of the 329 GPCRs we identified in *O. bimaculoides*. The full complement of GPCRs is presented in Supplementary Note 8.5. **a, b,** As reported for other lophotrochozoan genomes, the octopus genome contains chemosensory-like GPCRs; 74 GPCRs are similar to

the *Aplysia* chemosensory GPCRs⁵⁷ and 11 GPCRs are similar to vertebrate olfactory receptors. **c,** We identified 4 opsins in the octopus genome (from top to bottom): rhodopsin, rhabdomeric opsin, peropsin, and retinochrome. **d,** The octopus class F GPCRs comprises 6 genes: 5 Frizzled genes and 1 Smoothed gene (*). **e,** Thirty octopus genes show similarity to vertebrate adhesion GPCRs.



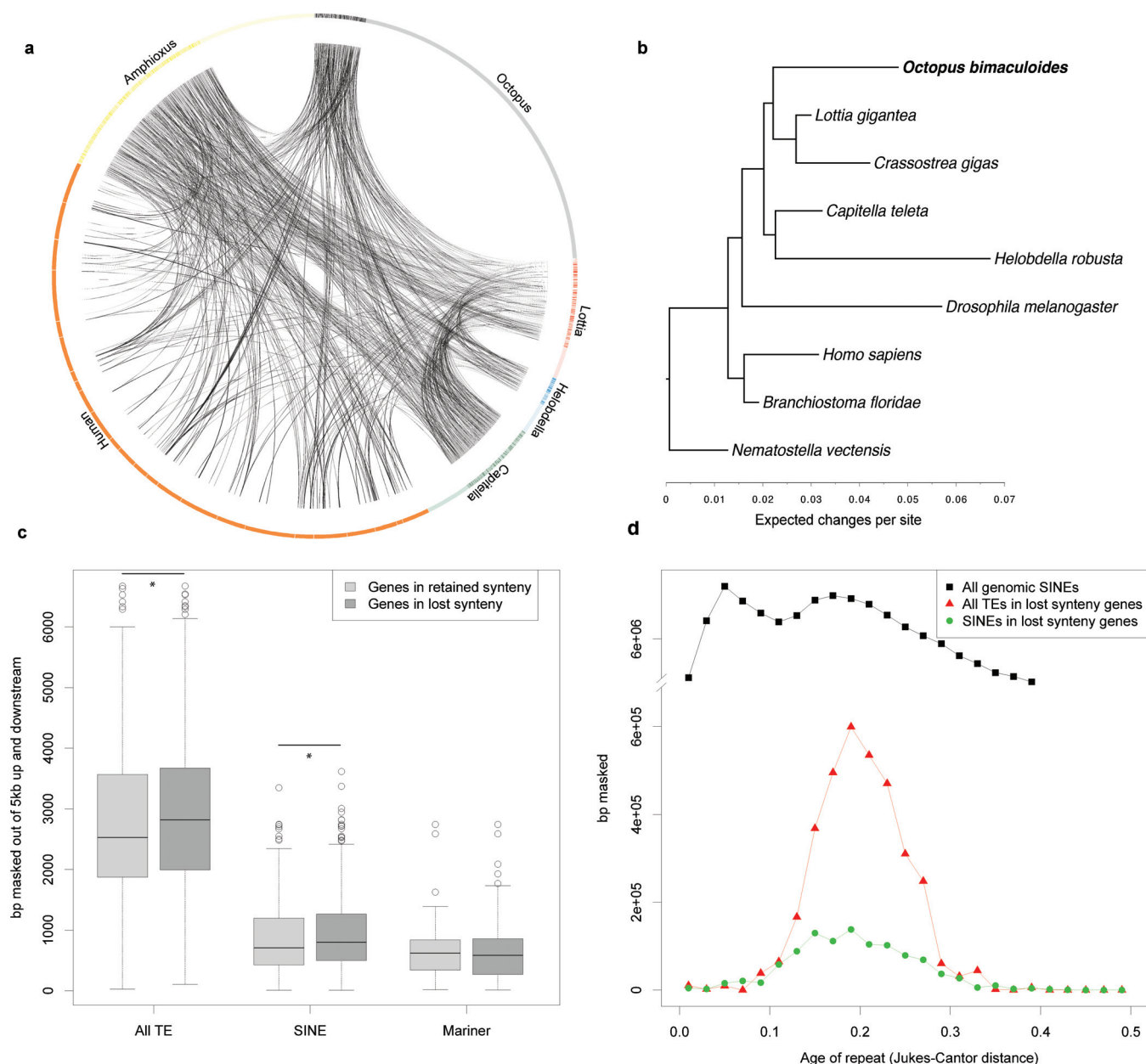
Extended Data Figure 7 | *O. bimaculoides* acetylcholine receptor (AChR) subunits. **a**, Phylogenetic tree of AChR subunit genes identified in Hsa, Mmu, Dme, Cte, Lgi and Obi. Black asterisk indicates a Dme sequence that groups with alpha 1-4-like subunits despite lacking two defining cysteine residues. **b**, Expression profiles of octopus AChR subunits. Genes ordered as in the tree (**a**), starting from the grey arrow and continuing anticlockwise. Putative non-ACh-binding subunits are highly expressed in the suckers. One sequence was not detected in our transcriptome data sets. In **a** and **b**, red asterisks indicate subunits with the substitution known to confer anionic permissivity⁵⁸.

c, Divergent octopus subunits lack nearly all residues necessary for ACh binding. Alignment of sequence flanking the cysteine loop (yellow) of the *L. stagnalis* ACh-binding protein (Lst_AchBP), the human and octopus alpha-7 receptor subunits (Hsa_AchR7, Obi_10697+), and the 23 divergent AChR subunits. Essential ACh-binding residues on the primary (pink) and complementary (blue) side of the ligand-binding domain are indicated²⁶, with conservative substitutions in a lighter shade. Outside of the binding residues, residues shared between the alpha-7 subunits are shaded in light grey, with bold letters for conservative substitutions.



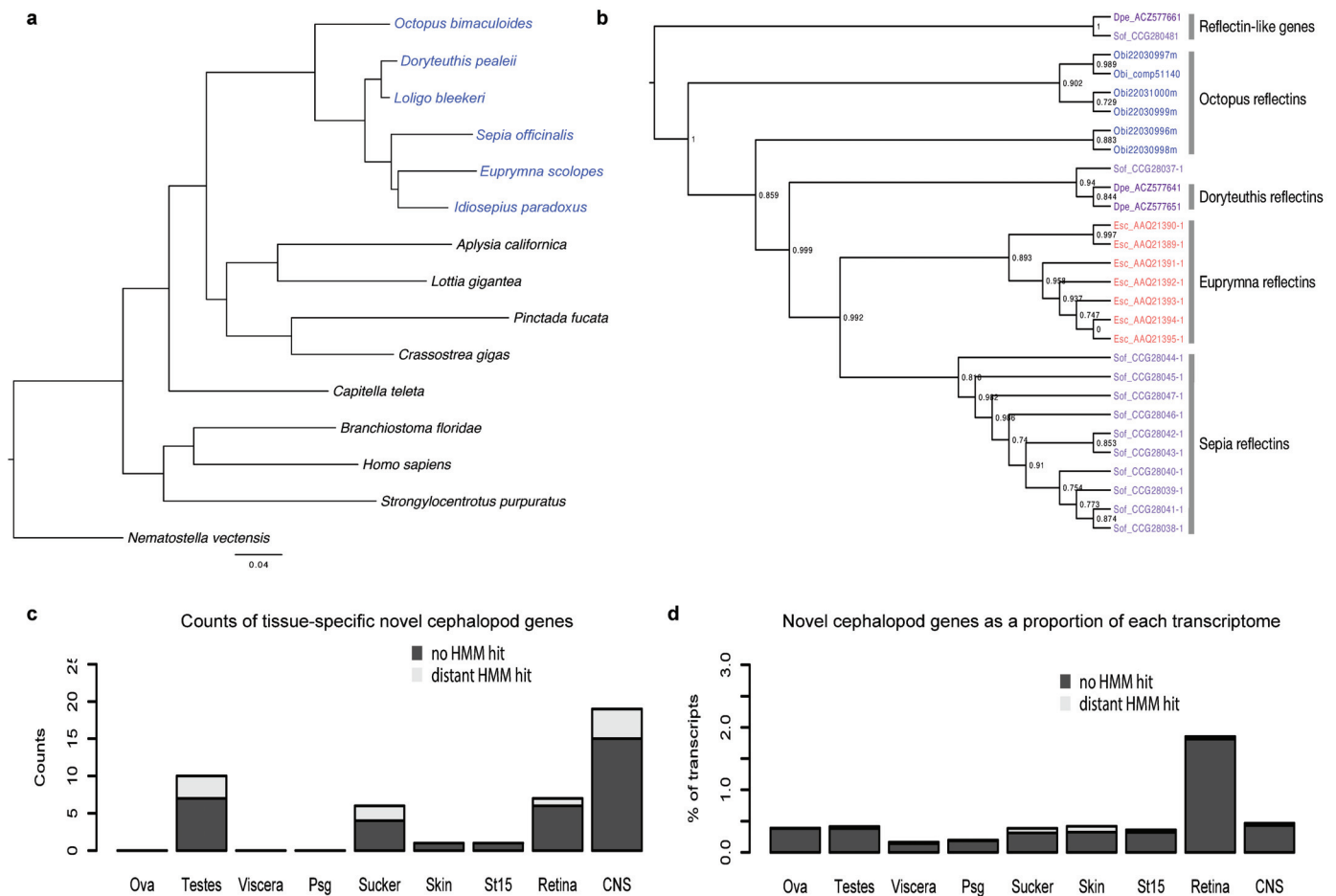
Extended Data Figure 8 | Active transposable elements and gene expression specificity. **a**, Transposable element expression across 12 tissues. **b**, Correlation between the total transposable element (TE) load (in bp) in the 5 kb regions flanking the gene and the fraction of genes with tissue-specific

expression (defined as having at least 75% of expression in a single tissue; see Source Data file for this figure). P value indicates the F -statistic for the significance of linear regression ($H_0: r^2 = 0$), with tissues with a P value ≤ 0.05 indicated in pink.



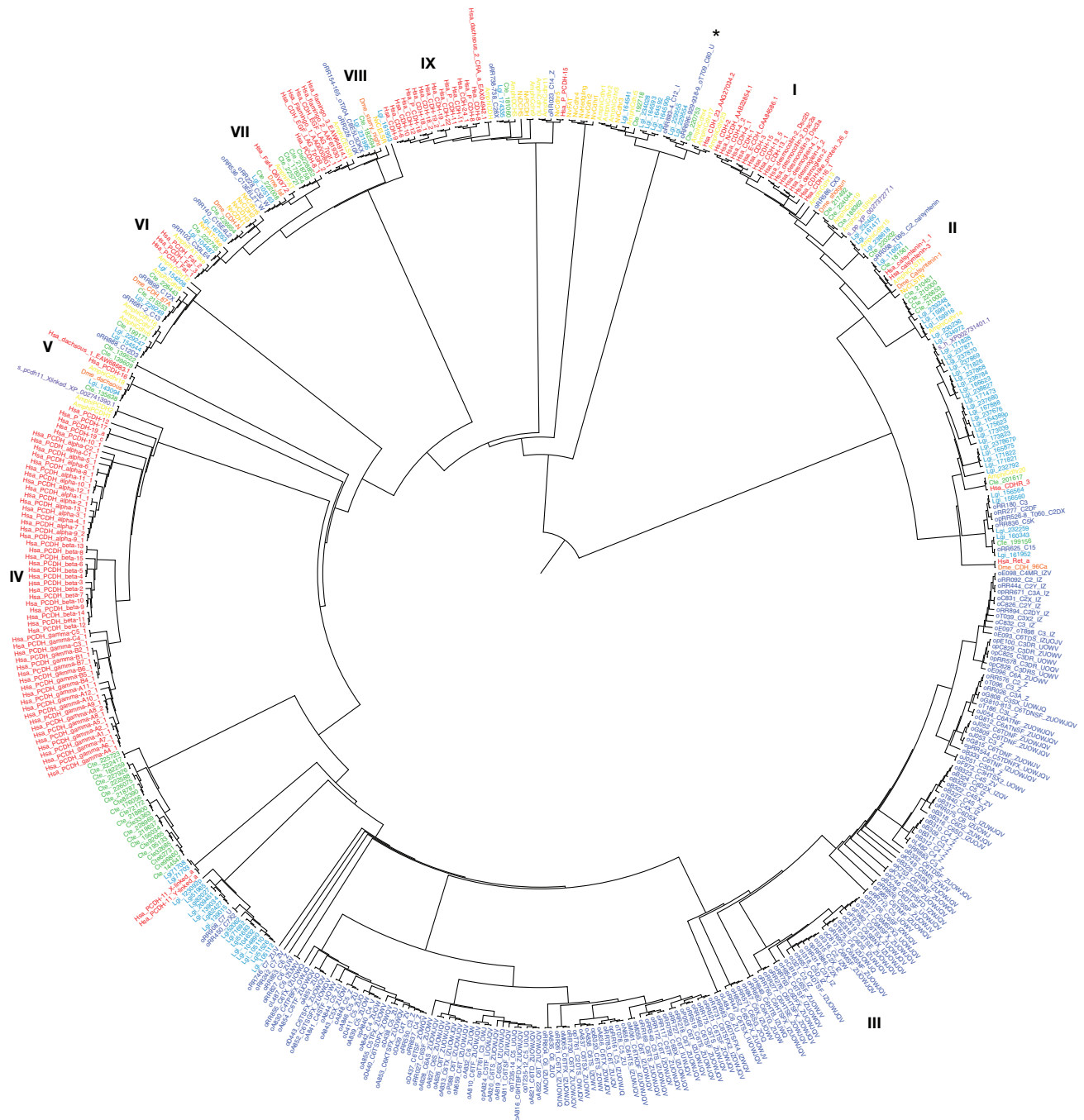
Extended Data Figure 9 | Synteny dynamics in octopus and the effect of transposable element (TE) expansions. **a**, Circos plot showing shared synteny across 6 genomes. Individual scaffolds are plotted according to bp length; scaffolds with no synteny are merged together (lighter arcs). Despite the large size of the octopus genome, only a small proportion of the scaffolds show synteny. **b**, Synteny reduction in octopus quantified based on synteny inference using gene families with at least one representative in human, amphioxus, *Capitella*, *Helobdella*, *Octopus*, *Lottia*, *Crassostrea*, *Drosophila*, and *Nematostella*. *Drosophila*, *Helobdella* and *Octopus* show the highest synteny

loss rates. Branch lengths, estimated with MrBayes⁵⁵, reflect extent of local genome rearrangement (Supplementary Note 6). **c**, Enrichment of overall and specific TE classes (base pairs masked) around genes from ancient bilaterian synteny blocks, including those absent in octopus (see key). Asterisks indicate Mann–Whitney *U*-test with *P* value < 0.02. **d**, Transposable element insertion history (Jukes–Cantor distance adjusted, see text) into the vicinity of genes from ‘lost’ synteny blocks. Note that only one SINE peak is present; a more recent peak (visible in ‘All genomic SINEs’) cannot be recovered from those insertions.



Extended Data Figure 10 | Cephalopod phylogeny and novelties. **a**, Whole-genome-derived phylogeny of molluscs and select other phyla showing the relative position of octopus at the base of the coleoid cephalopods. For methods see Supplementary Note 7.1. Members of the cephalopod class are indicated in blue, scale indicates number of substitutions per site. **b**, Phylogenetic tree of reflectin genes. Reflectins are cephalopod-specific genes that allow for rapid and reversible changes in iridescence. Six reflectin genes were identified in the

octopus genome. **c**, **d**, Novel gene expression across multiple tissues. Bars depict all cephalopod novelties; dark grey indicates sequences with no similarity to non-cephalopod genes using HMM searches (see Source Data for this figure). **c**, Counts of tissue-specific novelties in a given tissue. **d**, Proportion of expression of novel genes versus total expression in individual tissues. CNS (central nervous system) combines Supra, Sub, OL and ANC expression data.



Extended Data Figure 11 | Phylogenetic tree of cadherin genes. This is a larger image of Fig. 2a. Phylogenetic tree of cadherin genes in Hsa (red), Dme (orange), *Nematostella vectensis* (mustard yellow), *Amphimedon queenslandica* (yellow), Cte (green), Lgi (teal), Obi (blue), and *Saccoglossus kowalevskii* (purple). I, Type I classical cadherins; II, calsyntenins; III, octopus

protocadherin expansion (168 genes); IV, human protocadherin expansion (58 genes); V, dachous; VI, fat-like; VII, fat; VIII, CELSR; IX, Type II classical cadherins. Asterisk denotes a novel cadherin with over 80 extracellular cadherin domains found in Obi and Cte.

Identification of *cis*-suppression of human disease mutations by comparative genomics

Daniel M. Jordan^{1*}, Stephan G. Frangakis^{2*}, Christelle Golzio², Christopher A. Cassa¹, Joanne Kurtzberg³, Task Force for Neonatal Genomics[†], Erica E. Davis², Shamil R. Sunyaev^{1§} & Nicholas Katsanis^{2§}

Patterns of amino acid conservation have served as a tool for understanding protein evolution¹. The same principles have also found broad application in human genomics, driven by the need to interpret the pathogenic potential of variants in patients². Here we performed a systematic comparative genomics analysis of human disease-causing missense variants. We found that an appreciable fraction of disease-causing alleles are fixed in the genomes of other species, suggesting a role for genomic context. We developed a model of genetic interactions that predicts most of these to be simple pairwise compensations. Functional testing of this model on two known human disease genes^{3,4} revealed discrete *cis* amino acid residues that, although benign on their own, could rescue the human mutations *in vivo*. This approach was also applied to *ab initio* gene discovery to support the identification of a *de novo* disease driver in *BTG2* that is subject to protective *cis*-modification in more than 50 species. Finally, on the basis of our data and models, we developed a computational tool to predict candidate residues subject to compensation. Taken together, our data highlight the importance of *cis*-genomic context as a contributor to protein evolution; they provide an insight into the complexity of allele effect on phenotype; and they are likely to assist methods for predicting allele pathogenicity^{5,6}.

Understanding the nature and prevalence of genetic interactions has the potential to elucidate the evolutionary forces that act on protein residues, protein complexes and, more broadly, genomes. Some studies have reported that interactions are ubiquitous and contribute considerably to the evolutionary landscape^{7,8}, while others found that interactions are rare⁹. Even among those who agree that genetic interactions are important, the architecture of these interactions remains unclear: some studies find distinct interactions between two or three sites^{10,11}, while others propose a complex interaction network, effectively responding to aggregate properties of the entire protein or the entire genome^{12,13}.

One practical utility of comparative genomics has been highlighted by our appreciation of the large number of rare variants in humans and the difficulty in inferring their contribution to disease². To prioritize variants of interest, frequency in control populations and evolutionary conservation have become two prominent filters. Conserved regions are considered more likely to be intolerant of variation¹; programs such as PolyPhen⁵ and SIFT⁶ have employed this principle to predict functional effects of variants². Although useful, these strategies are constrained, in part because they do not take into consideration the genomic context of the mutated allele. An allele can appear damaging in one sequence yet be neutral in an orthologous sequence of another species. This phenomenon, referred to as compensated pathogenic deviation (CPD), contributes an unknown, but potentially large, number of false negatives to the evaluation of functional sites^{14,15}.

To examine the prevalence of CPDs, and to identify such sites, we used comparative genomics. A typical, non-CPD allele should cause the same phenotype in any orthologous sequence, regardless of genetic background. By contrast, when a variant that causes human genetic disease is found in a wild-type orthologous sequence, it is likely that the genetic background of that species exerts a compensatory effect on such a variant: it suppresses the phenotype, and thus protects the variant from negative selection^{11,14–16}. Previous studies have used this insight to quantify the fraction of CPD interspecies substitutions at ~10% (refs 14–16). Other studies have reported estimates of the inverse value, namely the fraction of pathogenic variants that are present as CPDs in other species, ranging from 2–18% (refs 17, 18). We set out to produce a new estimate of this value. We collected two data sets

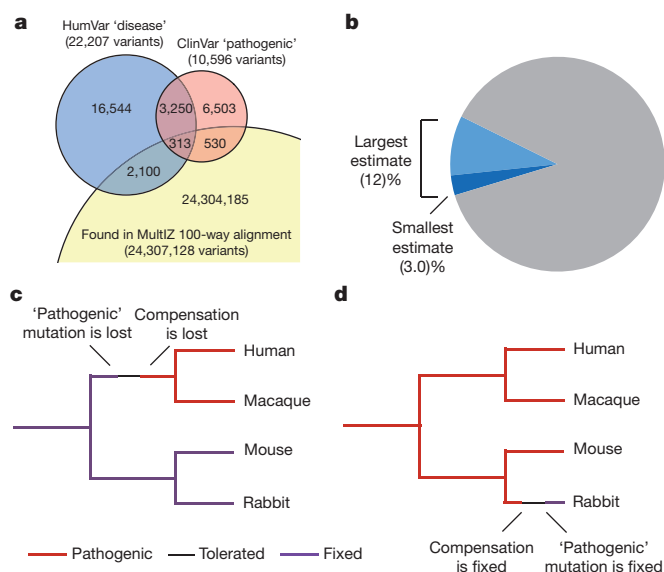


Figure 1 | Distribution of variants found in sequence alignments. **a**, Venn diagram showing sizes and overlap of the ClinVar and HumVar data sets, and how many are found in the multiple sequence alignment. **b**, Estimated number of human disease variants found in the alignment. The smallest estimate (3.0%, dark blue) comes from using the intersection of both variant data sets, requiring the variant to be absent from 6,503 human exomes, and filtering out alignments with low-quality scores. With any methodology, at least 88% of human disease variants (grey) are not found in the alignment. **c**, **d**, Potential mechanisms for the occurrence of CPDs in evolution. Branches where the variant is fixed are purple; branches where the variant is pathogenic are red. In **c**, the variant is present neutrally in an ancestor, but is lost in primates. Subsequent substitutions cause the ancestral allele to become pathogenic. In **d**, the variant is pathogenic in the ancestor, but mutations in a non-human branch cause it to become tolerated, and it arises later by mutation and becomes fixed.

¹Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. ²Center for Human Disease Modeling, Duke University, Durham, North Carolina 27701, USA. ³Department of Pediatrics, Division of Pediatric Blood and Marrow Transplantation, Duke University, Durham, North Carolina 27710, USA.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

[†]Lists of participants and their affiliations appear in the Supplementary Information.

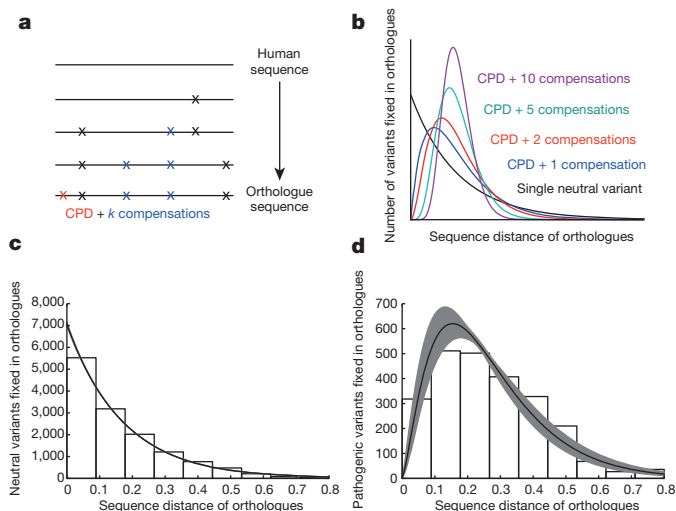


Figure 2 | Relationship between variants and evolutionary distance.

a, Model for fixation of CPDs. Neutral changes (crosses) arise neutrally. Some of these (blue) compensate for alleles that would otherwise be pathogenic. The parameter k represents the number of compensatory changes required for each pathogenic allele. Once k compensatory changes have fixed, the CPD (red) fixes neutrally. **b**, The relationship between evolutionary distance and the number of variants in the alignment is expected to be different for individual benign variants (black) and pathogenic variants with different numbers of compensations (blue, one; red, two; cyan, five; magenta, ten). **c**, **d**, Observed distribution of missense variants annotated as neutral (**c**) or pathogenic (**d**) in HumVar and present in vertebrate orthologues (bars), with maximum likelihood fits (black lines) and 95% confidence bands (grey shading). Panel **d** corresponds to a fitted value for k of 1.44 ± 0.07 .

of missense single-nucleotide variants (SNVs), annotated as either benign or pathogenic, derived from two databases, one based on the literature ('HumVar')^{5,19,20} and one based on clinical genetic laboratories and investigator submissions ('ClinVar')²⁰. Although the two databases are not fully independent, the majority of pathogenic variants were listed in one or the other (Fig. 1a). Overall, these data sets comprised 69,905 human missense mutations across 13,040 genes. We compared this data set to orthologous proteins from 100 vertebrates. As expected, we found the mutant residue for a large number of likely neutral human variants to be fixed in orthologues. However, the number of pathogenic missense variants found in orthologues (CPDs) was surprisingly high: $5.6\% \pm 0.5\%$ of ClinVar variants and $6.7\% \pm 0.4\%$ of HumVar variants were found in the alignment of mammals. For all vertebrates, these numbers increase to $10.2\% \pm 0.7\%$ and $12.0\% \pm 0.5\%$, respectively (Table 1).

Mindful of the possibility of false pathogenic annotations, we applied several filtering steps, including cross-referencing HumVar and ClinVar variant annotations with population frequency data, filtering on the basis of alignment quality²¹, using alternative alignment methodologies, and requiring variants to be present in multiple species (Table 1 and Supplementary Note). Some filters did remove bona fide recessive alleles (since we did not allow carriers), as well as disease variants with incomplete penetrance, even though this class of alleles is, by definition, sensitive to genomic context and thus likely to be affected by compensation²². Nevertheless, all filtering steps retained

a substantial number of variants (Supplementary Tables 1 and 2). Including only variants that pass all filtering steps and are detected in >1 vertebrate, we predict that the minimum estimate of CPDs in human patients is 3% (Fig. 1b). This is consistent with previous analyses, which have found that stringent filtering does not change the observed properties of CPDs to any notable extent^{16,17}. As a final test, we extracted *post hoc* pathogenic alleles from three different sources, each of which used independent means for assessing pathogenicity in acute paediatric disorders; overall, CPD rates ranged once again between 3% and 9%; additional analyses of other possible sources of bias were likewise consistent with our initial observations and previous studies (see Supplementary Note and Supplementary Tables 3, 4, 5).

We next turned to the question of the structure of the genetic interactions underlying such sites. Broadly, there are two possibilities: suppression of the disease phenotype may be the result of a small number of discrete compensatory substitutions; or suppression may be caused by a global shift in the properties of the gene, or the whole genome, caused by numerous substitutions that, individually, have small effects. The difference between these two models should be visible in the distribution of CPDs among orthologous sequences. During evolution, variants arise stochastically through a Poisson process: the expected amount of evolutionary time required to produce a given substitution is distributed exponentially²³. For a CPD, the distribution should be different; the presence of a CPD mandates the presence of all compensatory substitutions necessary for the CPD to be rendered neutral. As such, the expected evolutionary time required to produce a CPD is the sum of the times required to produce each compensatory substitution, followed by the time required to produce the CPD.

Previous studies have proposed different processes by which CPDs arise. The most plausible option is a neutral mechanism, in which the compensatory substitutions are neutral and arise/fix neutrally before the pathogenic substitution appears (Fig. 1c, d and Fig. 2a). In this case, the time required for each substitution to arise is given by an exponential distribution, and the time for all compensatory sites to arise is approximated by the convolution of multiple exponential distributions (a gamma distribution, in the case where all exponential distributions are identical). The number of exponential distributions included in the convolution corresponds to one plus the number of compensatory substitutions required, and it can be inferred from the shape of the distribution (Fig. 2b).

Although the evolutionary time separating two sequences is not observable directly, we can approximate it using sequence distance (one minus sequence identity)²⁴. We plotted the number of missense variants observed as a function of sequence distance for neutral variants and for CPDs. Qualitatively, the shapes of both distributions match theoretical expectations. The two distributions are distinct from each other ($P = 1.6 \times 10^{-68}$, Kolmogorov–Smirnov two-sample test; Supplementary Tables 6, 7). Additionally, the observed distribution of CPDs is weighted towards shorter evolutionary distances, as expected if most CPDs require a small number of individual compensatory substitutions, as opposed to the normal distribution expected if CPDs require many individual compensatory substitutions (Fig. 2b, d). To obtain a more precise estimate of the number of compensatory substitutions, we used maximum likelihood to fit several versions of the convolution-of-exponentials model with different combinations of variant data sets and alignment strategies (Fig. 2c, d; see Methods

Table 1 | Range of estimates for prevalence of CPDs in human disease

	Unfiltered MultiZ alignment	High-quality MultiZ alignments	Mammalian subset of MultiZ alignment	Present in >1 species in MultiZ alignment	EPO alignment
HumVar	$12.0\% \pm 0.5\%$	$11.5\% \pm 0.5\%$	$6.7\% \pm 0.4\%$	$6.1\% \pm 0.3\%$	$7.5\% \pm 0.4\%$
ClinVar	$10.2\% \pm 0.7\%$	$9.9\% \pm 0.7\%$	$5.6\% \pm 0.5\%$	$4.7\% \pm 0.5\%$	$6.5\% \pm 0.6\%$
HumVar+ClinVar	$9.3\% \pm 1.0\%$	$8.5\% \pm 1.0\%$	$5.3\% \pm 0.8\%$	$3.9\% \pm 0.7\%$	$5.5\% \pm 0.9\%$
HumVar+ClinVar+ESP	$7.5\% \pm 1.0\%$	$7.0\% \pm 1.0\%$	$3.8\% \pm 0.7\%$	$3.0\% \pm 0.6\%$	$4.0\% \pm 0.8\%$

Fraction of likely pathogenic mutations in humans considered to be CPDs according to different filtering paradigms. Values represent the fraction of variants for which an alignment could be retrieved where the variant amino acid is present in an orthologue sequence; error ranges are Jeffreys 95% confidence intervals. ESP, NHLBI Exome Sequencing Project; EPO, Enredo–Pecan–Ortheus pipeline.

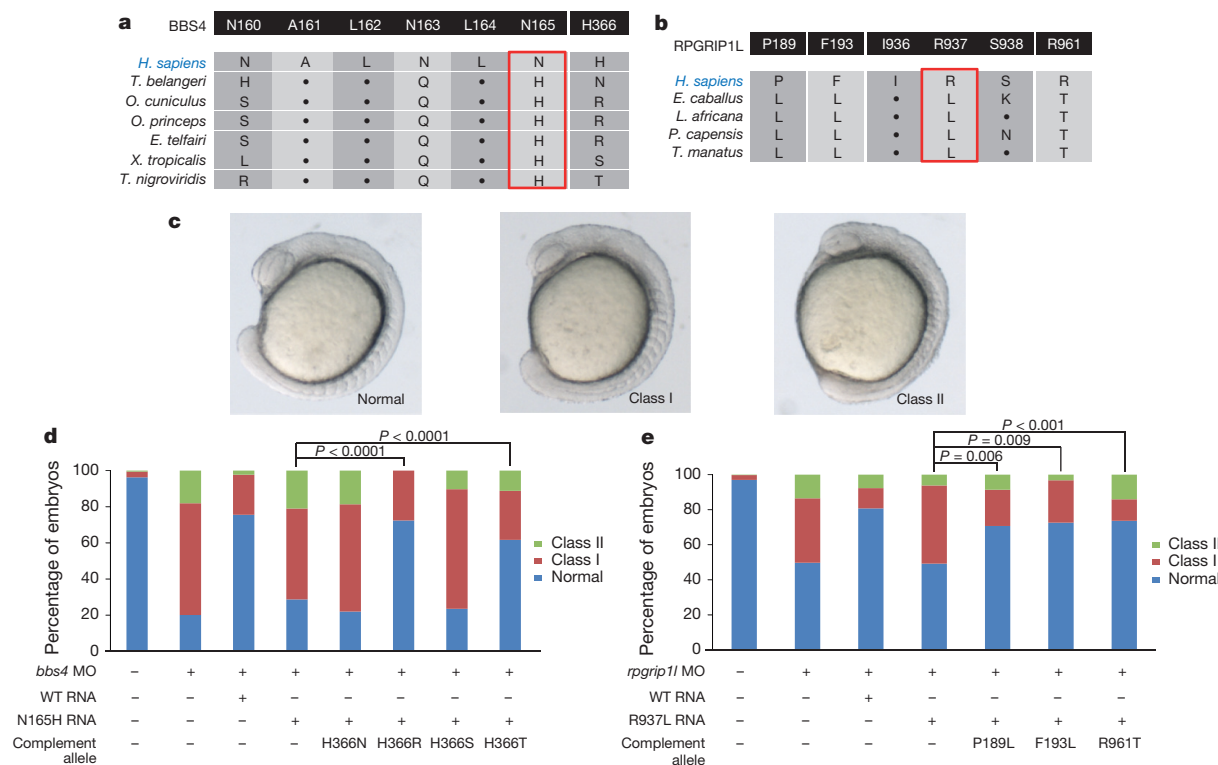


Figure 3 | Compensatory mutations rescue pathogenic alleles in *BBS4* and *RPGRIP1L*. **a**, The pathogenic BBS4 165H allele is fixed in six species. Secondary sites 160, 163 and 366 are possible CPDs. **b**, The pathogenic RPGRIP1L 937L allele is fixed in four species. The 189L, 193L and 961T alleles are present in all four species. **c**, Examples of zebrafish convergent extension phenotypic groups. **d**, Human RNA encoding the BBS4 165H mutation and

either 366R or 366T can rescue the morphant phenotype; RNA encoding 165H mutation alone cannot. WT, wild type. **e**, Mutation of 189L, 193L or 961T, in the background of 937L *RPGRIP1L* mRNA, rescues the loss of function observed in 937L RNA. Significance was determined by χ^2 test. See Supplementary Table 9 for embryo counts.

and Supplementary Tables 6, 7). Most versions of the model fit best as the convolution of approximately two exponential distributions, supporting a mechanism in which most CPDs are compensated by simple pairwise interactions. Additionally, most models reported similar rates of evolution for neutral variants, CPDs and compensatory variants, suggesting that the target size for compensatory changes is small. We repeated these analyses with multiple different variant data sets and alignment strategies, finding similar results each time (Extended Data Fig. 1 and Supplementary Table 8).

These analyses predict that most CPDs could be rescued by one large-effect compensatory substitution. We tested this prediction experimentally. We posited that each vertebrate sequence that includes a CPD should also include its *cis*-compensatory allele. Therefore, every amino acid difference between the human sequence and the sequence of the orthologue(s) containing a CPD is a candidate compensatory substitution. Given the practical constraints of examining all possible compensatory substitutions in macromolecular complexes, we focused on substitutions within the same gene as the CPD.

Scanning our list of candidate CPDs, we noted two alleles in genes involved in ciliopathies: a protein-encoding p.N165H change in *BBS4* and a p.R937L variant in *RPGRIP1L*, which contribute pathogenic alleles to Bardet–Biedl syndrome and Meckel–Gruber syndrome, respectively^{3,4}. These alleles were prioritized because: (1) Bardet–Biedl and Meckel–Gruber syndromes have a severe effect on reproductive fitness; (2) previous studies have established loss-of-function zebrafish phenotypes rescuable by human messenger RNA for both genes^{4,25}; (3) *in vivo* complementation has indicated both alleles to be deleterious to human protein function^{4,25}; and (4) we observed multiple species with the human mutant allele fixed: six species for BBS4 165H and four for RPGRIP1L 937L (Fig. 3a, b)—for this reason, both alleles were predicted to be benign (PolyPhen-2, SIFT, MutationAssessor).

Comparative genomic analysis identified 9 candidate sites in BBS4 and 32 candidate sites in RPGRIP1L (Supplementary Table 9). To test each site, we took advantage of the established convergent extension defects induced by morpholino (MO)-mediated suppression of *bbs4* or *rpgr11* in zebrafish^{4,25}. Consistent with previous observations, suppression of *bbs4* or *rpgr11* induced convergent extension defects in 80% and 50% of embryos respectively ($n = 50$ –100 embryos; Fig. 3c–e). Co-injection of MO with human wild-type mRNA rescued this phenotype, whereas injection with human mutant mRNA showed no improvement (Fig. 3d, e). We next tested the entire candidate complementing allelic series for each gene. For *BBS4*, the introduction of 2/9 candidate residues in *cis* with the 165H-encoding mRNA ameliorated the phenotype in a manner indistinguishable from wild-type mRNA. Strikingly, both complementing alleles affected the same amino acid and were specific to the compensatory changes: the 165H/366N and the 165H/366S behaved as null, whereas 165H/366R was indistinguishable from wild type; 165H/366T converted the functional null to a hypomorph (Fig. 3d and Extended Data Fig. 2a).

We observed a similar pattern for *RPGRIP1L*. Testing each of the 32 candidate sites identified three complementing events, two of which map to the same region: 937L/189L, 937L/193L and 937L/961T (Fig. 3e and Extended Data Fig. 2b). Testing each complementing allele individually showed them to be either extremely mild or benign (Supplementary Table 9). Finally, comparative genomic analysis showed that these data could explain the tolerance of the RPGRIP1L 937L change in all four species and of the BBS4 165H change in 4/6 species (Fig. 3a, b).

The above analysis is limited by its retrospective nature. We therefore tested the usefulness of our model in *ab initio* gene discovery. We have recently initiated a whole-exome sequencing (WES) and func-

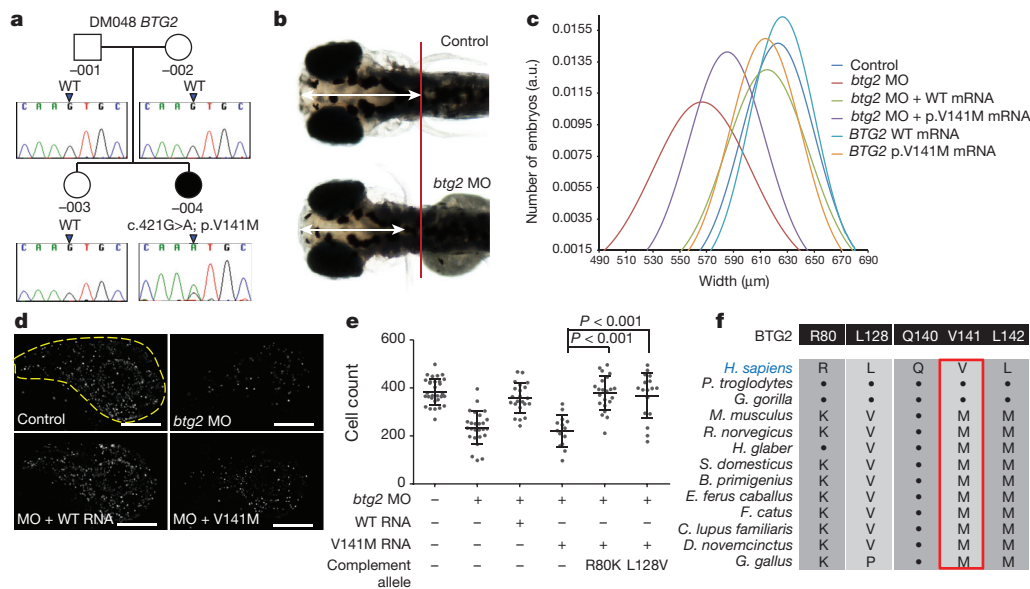


Figure 4 | A *de novo* *BTG2* p.V141M-encoding allele causes microcephaly. **a**, Pedigree DM048. Chromatograms show a *de novo* c.421G>A nucleotide change. WT, wild type. **b**, Suppression of *btg2* leads to head size defects. Dorsal view of uninjected control and *btg2* MO-injected zebrafish embryos at 4 dpf. White arrows show the distance measured from forebrain to hindbrain. Red line shows the protrusion of the pectoral fins in uninjected controls. **c**, Distribution of head size measurements at 4 dpf (Supplementary Table 10;

tional testing paradigm to accelerate gene discovery in young children called Task Force for Neonatal Genomics (TFNG). Patients who display anatomical phenotypes amenable to functional modelling in zebrafish are evaluated by trio-based WES and have candidate alleles tested systematically *in vivo*²⁶.

We enrolled a 17-month-old female with an undiagnosed neuro-anatomical condition hallmarked by microcephaly (Fig. 4a). We filtered WES data for non-synonymous and splice variants with a minor allele frequency of <1%, and we conducted a proband-centric trio analysis that yielded four candidates: *de novo* missense changes in *BTG2* and *NOS2*; and recessive missense variants in *TTN* and *LAMA1*. Testing of an unaffected sibling excluded *LAMA1*; *TTN*, a known dominant cardiomyopathy locus²⁷, is an unlikely driver.

To investigate the pathogenicity of the *BTG2* (p.V141M) and *NOS2* (p.P795A) protein-encoding changes, we studied *btg2* and *nos2* in zebrafish. Reciprocal use of Basic Local Alignment Search Tool (BLAST) between *Homo sapiens* and *Danio rerio* identified a single zebrafish *btg2* orthologue and two zebrafish *nos2* orthologues. We injected splice-blocking MO (sb-MO) or translational-blocking MO (tb-MO) (Extended Data Fig. 3) into zebrafish embryos (3 ng; *n* = 80 embryos per injection) and scored for head size defects at 4 days post-fertilization (dpf) by measuring the anterior–posterior distance between the forebrain and the hindbrain (Fig. 4b). For *nos2a/b* MO-injected embryos, we saw no differences at the highest dose injected (8 ng for *nos2a/b* sb-MOs; Supplementary Table 10). By contrast, we found a significant reduction of anterior structures in *btg2* morphants ($P < 0.0001$; Fig. 4b, c). Co-injection of wild-type human *BTG2* mRNA with tb-MO resulted in significant rescue ($P < 0.0001$; Fig. 4c). In contrast, injection of mRNA harbouring 141M was significantly worse at rescue than wild type ($P < 0.0001$; Fig. 4c).

BTG2 is a regulator of cell cycle checkpoint in neuronal cells²⁸ and is strikingly intolerant to variation in humans (Exome Variant Server (EVS)). To test the pathogenicity of 141M by a different assay, we performed antibody staining at 2 dpf (a time before the manifestation of microcephaly). We marked post-mitotic neurons in the forebrain with antibodies against neuronal HuC/HuD antigens, and we scored (blind, triplicate) on the basis of an established paradigm²⁹. *btg2*

white arrows in **b**), a.u., arbitrary units. **d**, 2 dpf zebrafish embryos stained for PH3. Human RNA containing the V141M mutation is unable to rescue the reduced proliferation of *btg2* morphants. Scale bars, 250 μ m. **e**, Quantification of PH3-positive cells: human RNA with mutations V141M and either R80K or L128V can rescue knockdown of *btg2*. Error bars represent standard deviation. **f**, The 141M allele is fixed in 59/87 species besides primates, examples displayed here. See Supplementary Table 11 for PH3 quantification.

morphants displayed a significant decrease in HuC/HuD staining ($P < 0.0001$; Extended Data Fig. 4). This defect was rescued with wild-type *BTG2* mRNA ($P < 0.05$), but could not be ameliorated by 141M-encoded mRNA co-injection (Extended Data Fig. 4). Importantly, co-injection of *btg2* tb-MO with two rare control EVS alleles (p.A126S and p.R145Q) resulted in rescue, providing evidence for assay specificity (Extended Data Fig. 4b). As a third test, we stained whole embryos with a phospho-histone H3 (PH3) antibody that marks proliferating cells. We counted the number of positive cells in a defined anterior region of embryos. We saw a significant reduction in cell proliferation in the heads of 2 dpf *btg2* morphants ($P < 0.0001$); this defect was likewise rescued by co-injection of wild-type mRNA, while 141M mutant rescue was indistinguishable from *btg2* tb-MO alone ($P = 0.38$; Fig. 4b, d). Combined, all three assays indicated that *BTG2* p.V141M is pathogenic and that haploinsufficiency of this gene probably contributes to the microcephaly of the proband.

Despite our functional and genetic data for p.V141M, this allele was predicted computationally to be benign. A likely reason is that, with the exception of primates, most *BTG2* orthologues encode Met at the orthologous position (Fig. 4f). These data suggested that V141 might represent a CPD site in primates that branched from the ancestral methionine. To test this possibility, we identified nine *BTG2* sites that co-evolved with 141M (Supplementary Table 11), which we mutagenized into the human construct encoding 141M. We then injected embryos with *btg2* MO; MO plus wild-type human *BTG2* mRNA; MO plus 141M-encoding mRNA; or MO plus 141M in *cis* with one of the nine candidate complementing alleles. Seven of the alleles had no effect (Supplementary Table 11). However, R80K- or L128V-encoded mRNA on the 141M backbone rescued the number of PH3-positive cells to wild-type levels (Fig. 4e and Extended Data Fig. 2c); both alleles were benign on their own (Supplementary Table 11). Taken together, our data indicated that 141M is deleterious in the human background, but the protection of this residue conferred by either Lys 80 or by Val 128 can explain >90% (54/59) of species encoding 141M (Fig. 4f).

To improve the scalability of detecting CPDs, we used our model of CPD evolution to develop a computational predictor for distinguish-

ing variants that are unlikely to be CPDs from those that might be CPDs, and to identify candidate compensations to aid experimental design (<http://genetics.bwh.harvard.edu/cpd/>). Initial testing of this tool intimated high negative predictive values but modest positive predictive values, probably due to the dearth of known CPDs (Supplementary Note).

Our results contrast with some previous studies that claim that epistasis is ubiquitous^{7,10}; or that it is practically nonexistent⁹; or that it is commonly of higher order^{12,13}. The most likely explanation for this discrepancy is that such studies have examined different kinds of variation and traits. For example, studies on the evolution of genetic incompatibilities rely on assumptions of high mutation rate and weak negative selection, assumptions that generally do not hold for the case of pathogenic missense variation^{10,30}. The difference with the studies suggesting higher-order *cis*-interactions may be to do with the scale of evolutionary time our analyses probe: the span of hundreds of millions of years of evolution represented by the vertebrate alignment may not be long enough to reveal higher-order combinations of non-synonymous SNVs. Indeed, using neutral SNVs from the HumVar data set as a control, we estimate the vertebrate alignment has explored 12% of pairwise interactions between SNVs, compared to 0.6% of three-way interactions between SNVs. It is possible that higher-order interactions are common, but are not detectable without a deeper alignment.

Finally, considering the accelerated use of genome editing to model human pathogenic mutations in a variety of model organisms, our data highlight the critical need to not only pair computational predictions with functional studies, but also to evaluate the effect of human mutations in the context of the human sequence.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 October 2014; accepted 23 April 2015.

Published online 29 June 2015.

- Alföldi, J. & Lindblad-Toh, K. Comparative genomics as a tool to understand evolution and disease. *Genome Res.* **23**, 1063–1068 (2013).
- Cooper, G. M. & Shendure, J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature Rev. Genet.* **12**, 628–640 (2011).
- Katsanis, N. *et al.* *BBS4* is a minor contributor to Bardet-Biedl syndrome and may also participate in triallelic inheritance. *Am. J. Hum. Genet.* **71**, 22–29 (2002).
- Khanna, H. *et al.* A common allele in *RPGRIP1L* is a modifier of retinal degeneration in ciliopathies. *Nature Genet.* **41**, 739–745 (2009).
- Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248–249 (2010).
- Sim, N. L. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, W452–W457 (2012).
- Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. A. Epistasis as the primary factor in molecular evolution. *Nature* **490**, 535–538 (2012).
- Weinreich, D. M., Delaney, N. F., DePristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
- McCandlish, D. M., Rajon, E., Shah, P., Ding, Y. & Plotkin, J. B. The role of epistasis in protein evolution. *Nature* **497**, E1–2 (2013).
- Corbett-Detig, R. B., Zhou, J., Clark, A. G., Hartl, D. L. & Ayroles, J. F. Genetic incompatibilities are widespread within species. *Nature* **504**, 135–137 (2013).
- Gao, L. & Zhang, J. Why are some human disease-associated mutations fixed in mice? *Trends Genet.* **19**, 678–681 (2003).
- Weinreich, D. M., Lan, Y., Wylie, C. S. & Heckendorn, R. B. Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev.* **23**, 700–707 (2013).
- Chou, H. H., Chiu, H. C., Delaney, N. F., Segre, D. & Marx, C. J. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011).
- Kondrashov, A. S., Sunyaev, S. & Kondrashov, F. A. Dobzhansky-Muller incompatibilities in protein evolution. *Proc. Natl Acad. Sci. USA* **99**, 14878–14883 (2002).
- Kulathinal, R. J., Bettencourt, B. R. & Hartl, D. L. Compensated deleterious mutations in insect genomes. *Science* **306**, 1553–1554 (2004).
- Soylemez, O. & Kondrashov, F. A. Estimating the rate of irreversibility in protein evolution. *Genome Biol. Evol.* **4**, 1213–1222 (2012).
- Ferrer-Costa, C., Orozco, M. & de la Cruz, X. Characterization of compensated mutations in terms of structural and physico-chemical properties. *J. Mol. Biol.* **365**, 249–256 (2007).
- Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Mottaz, A., David, F. P., Vuthey, A. L. & Yip, Y. L. Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. *Bioinformatics* **26**, 851–852 (2010).
- Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–D985 (2014).
- Ahola, V., Aittokallio, T., Vihinen, M. & Uusipaikka, E. Model-based prediction of sequence alignment quality. *Bioinformatics* **24**, 2165–2171 (2008).
- Giudicessi, J. R. & Ackerman, M. J. Determinants of incomplete penetrance and variable expressivity in heritable cardiac arrhythmia syndromes. *Transl. Res.* **161**, 1–14 (2013).
- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, 1984).
- Povolotskaya, I. S. & Kondrashov, F. A. Sequence space and the ongoing expansion of the protein universe. *Nature* **465**, 922–926 (2010).
- Zaghloul, N. A. *et al.* Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. *Proc. Natl Acad. Sci. USA* **107**, 10602–10607 (2010).
- Katsanis, N., Cotten, M. & Angrist, M. Exome and genome sequencing of neonates with neurodevelopmental disorders. *Future Neurology* **7**, 655–658 (2012).
- Herman, D. S. *et al.* Truncations of titin causing dilated cardiomyopathy. *N. Engl. J. Med.* **366**, 619–628 (2012).
- Montagnoli, A., Guardavaccaro, D., Starace, G. & Tirone, F. Overexpression of the nerve growth factor-inducible *PC3* immediate early gene is associated with growth inhibition. *Cell Growth Differ.* **7**, 1327–1336 (1996).
- Beunders, G. *et al.* Exonic deletions in *AUTS2* cause a syndromic form of intellectual disability and suggest a critical role for the C terminus. *Am. J. Hum. Genet.* **92**, 210–220 (2013).
- Fraisse, C., Elderfield, J. A. & Welch, J. J. The genetics of speciation: are complex incompatibilities easier to evolve? *J. Evol. Biol.* **27**, 688–699 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Y. Liu and D. Balick for helpful discussions, M. Kousi for assistance with the NCL mutational list, and M. Talkowski, A. Kondrashov and G. Lyon for critical review of the manuscript. This work was supported by grants R01HD04260, R01DK072301 and R01DK075972 (N.K.); R01 GM078598, R01 MH101244, R01 DK095721 and U01 HG006500 (S.R.S.); R01EY021872 (E.E.D.); and a NARSAD Young Investigator Award (C.G.). N.K. is a Distinguished Brumley Professor.

Author Contributions D.M.J., S.G.F., S.R.S. and N.K. designed the overall study. D.M.J., C.A.C. and S.R.S. conceptualized the principle of CPDs and performed all computational analyses. S.G.F., E.E.D. and N.K. conceptualized the biological properties of CPDs and implemented *in vivo* testing with the assistance of C.G. J.K. referred the index patient and evaluated clinical data in the context of molecular discoveries. The Task Force for Neonatal Genomics constructed the platforms and methods for recruitment, ascertainment and evaluation of clinical and molecular data and return of results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R.S. (ssunyaev@rics.bwh.harvard.edu) or N.K. (nicholas.katsanis@dm.duke.edu).

METHODS

Data sets of known benign and pathogenic variants. Our primary training data set was HumVar, one of the training data sets for PolyPhen-2.2.3. We used the most recent public release at the time of this publication (December 2011), available for download at <http://genetics.bwh.harvard.edu/pph2/dokuwiki/downloads>. This data set is derived from SwissVar variant annotations¹⁹. It contains 22,207 variants annotated as pathogenic and 21,433 variants annotated as benign. We also used a data set of pathogenic variants derived from the June 2014 release of the ClinVar database²⁰. This data set consists of all missense variants from ClinVar that are unambiguously (that is, classified the same by all submitters) and confidently (that is, not a 'Likely' annotation) annotated as 'Pathogenic'. It contains 10,596 variants annotated as pathogenic and 1,926 variants annotated as benign. The intersection of these two data sets contains 3,563 variants annotated as pathogenic and 454 variants annotated as benign. As an additional control, we required that the pathogenic variants be absent from 6,503 human exomes³¹ (EVS; <http://evs.gs.washington.edu/EVS/>). This most stringent data set contains 3,062 variants annotated as pathogenic.

Comparative genomics screen for CPDs. We used the University of California, Santa Cruz (UCSC) MultiZ whole-genome alignments of 100 vertebrate sequences³², downloaded from UCSC as translated exons. As an alternative alignment strategy, we used the EPO alignment of 37 eutherian mammal species³³, downloaded as nucleotide sequences and translated for all aligned species using the human open reading frame. In cases in which the alignment contained multiple sequences from the same species, only the sequence most similar to the human sequence was retained. Variants were classified as CPDs if the variant amino acid was found in the translated sequence of any vertebrate orthologue other than human or chimpanzee, with chimpanzee being excluded because presence in the chimpanzee sequence may be used as evidence for neutrality in variant annotation databases. The resulting data set of neutral and deleterious variants found in vertebrate orthologues is available (see Source Data for Fig. 1).

Statistical models of variant density. We modelled the density of benign variants with an exponential distribution, with scale parameter β_{neut} representing the mean time to fixation of neutral alleles. We used three different models for the density of CPDs. (1) k compensatory changes fix at rate $1/\theta$, followed by the now-neutral CPD at the same rate. This is represented by a gamma distribution with shape parameter $k + 1$ and scale parameter θ . (2) k compensatory changes fix at rate $1/\theta_1$, followed by the now-neutral CPD at an independent rate $1/\theta_2$. This is represented by the convolution of a gamma distribution with shape parameter k and scale parameter θ_1 , and an exponential distribution with scale parameter θ_2 . (3) k compensatory changes fix at rate $1/\beta_{\text{comp}}$, followed by the now-neutral CPD at the neutral rate $1/\beta_{\text{neut}}$.

We assume a reversible model of evolution, so that the same three models can apply both to the fixation of CPDs not present in an ancestral sequence and to the loss of CPDs that are present in an ancestral sequence. The random variable used in these models is the sequence distance to the closest sequence containing the variant, where sequence distance is defined as the fraction of aligned (that is, non-gapped) positions that are identical.

Fitting observed density to statistical models. We recorded for each variant found in a vertebrate orthologue the minimum number of amino acid differences between that variant and a vertebrate orthologue, not counting gapped sites, normalized by the length of the sequence. We then used maximum likelihood to fit the neutral model and each of the three pathogenic models described earlier, using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization algorithm to maximize the likelihood functions. We repeated the fit using each of our filtered variant data sets and alignment methodologies, as well as discarding all variants that were only found in the alignment in a single sequence. All three models fit reasonably well, and all produced qualitatively similar results on all data sets and alignment methodologies. The exact fitted parameter values are found in Supplementary Table 8.

Prediction method. Our prediction method is implemented in Perl, and the source code is available (Supplementary Data 1-CPD Predictor Code). To calculate the probability that a variant is a CPD, we find the minimum distance to the variant in the multiple sequence alignment and apply Bayes' Law. We use the third likelihood model described earlier, where the CPD fixes at the neutral rate, using

the maximum likelihood inferred parameter values. As our prior we use the well-established result that 10% of variants seen in another sequence are CPDs. Candidate compensation sites are identified by collecting all substitutions found in any sequence containing the candidate CPD, prioritizing sites that are substituted in many sequences over sites that are substituted in only a few sequences.

WES. Research study participants were enrolled upon informed consent according to protocols approved by the Duke University Internal Review Board. We conducted paired-end pre-capture library preparation by fragmenting genomic DNA through sonication, ligating it to the Illumina multiplexing PE adapters, and PCR amplification using indexing primers. For target enrichment/exome capture we enriched the pre-capture library by hybridizing to biotin-labelled VCRome 2.1 (ref. 34) in-solution Exome Probes at 47 °C for 64–72 h. For massively parallel sequencing, the post-capture library DNA was subjected to sequence analysis on an Illumina HiSeq platform for 100 bp paired-end reads (130× median coverage, >95% target coverage at 10×). Primary data were interpreted and analysed by Mercury 1.0; the output data from Illumina HiSeq were converted from bcl files to FastQ files by Illumina CASAVA 1.8 software, and mapped by the BWA program. We performed variant calls using Atlas-SNP and Atlas-indel³⁵.

Morpholino design. MOs targeting zebrafish *bbs4* and *rpgril1* were obtained from Gene Tools, and described previously^{4,25}. MOs against zebrafish *btg2* and *nos2* were obtained from Gene Tools (Extended Data Fig. 3; sequences available upon request).

Site-directed mutagenesis. Mutant alleles were generated as described²⁵. Sequences were validated via Sanger sequencing on Applied Biosystems 3730xl DNA Analyzer.

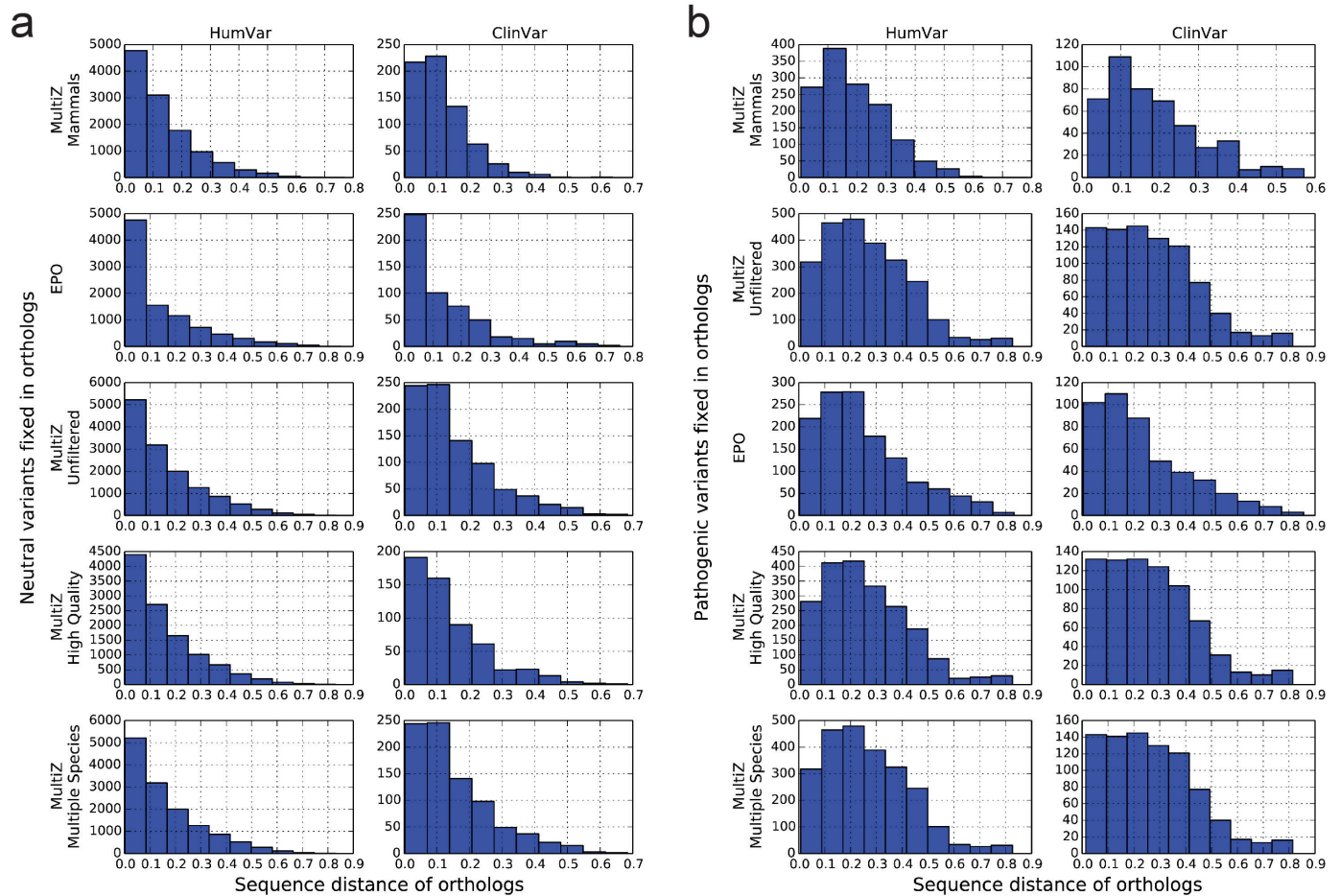
mRNA synthesis and zebrafish embryo injection. mRNA was transcribed *in vitro* as described³⁶ using SP6 Message Machine kit (Ambion). MO and mRNA concentrations were determined based on the combination by which wild-type mRNA efficiently rescued the morphant phenotype. The same concentrations were used for rescue with mutant mRNA or injection of mRNA alone. The MO and mRNA concentrations injected were as follows: 0.7 ng *bbs4* MO and 100 pg *BBS4* mRNA; 5 ng *rpgril1* MO and 100 pg *RPGRIP1L* mRNA; 3 ng *btg2* MO and 150 pg *BTG2* mRNA; 8 ng *nos2a*; 8 ng *nos2b*. All animal work was performed in accordance with the protocols and guidelines of the Duke Institutional Animal Care and Use Committee.

Classification and scoring of embryos. Embryos injected with *bbs4* or *rpgril1* MOs were classified into two graded phenotypes on the basis of the relative severity compared with age-matched uninjected controls from the same clutch, as described previously²⁵. Comparisons between injections of MO alone, mRNA alone, mutant rescue, and wild-type rescue were performed by χ^2 test.

Embryos injected with *btg2* MO at were fixed in 4% paraformaldehyde at either 2 dpf or 4 dpf. Two days post-fertilization embryos were stained for HuC/HuD or PH3. HuC/HuD was scored and quantified as described²⁵. Four days post-fertilization embryos were transferred to 1× PBS and bright-field dorsal images were captured; we assessed head size by measuring the distance from the anterior-most region of the forebrain to the hindbrain as defined by the attachment of the pectoral fins.

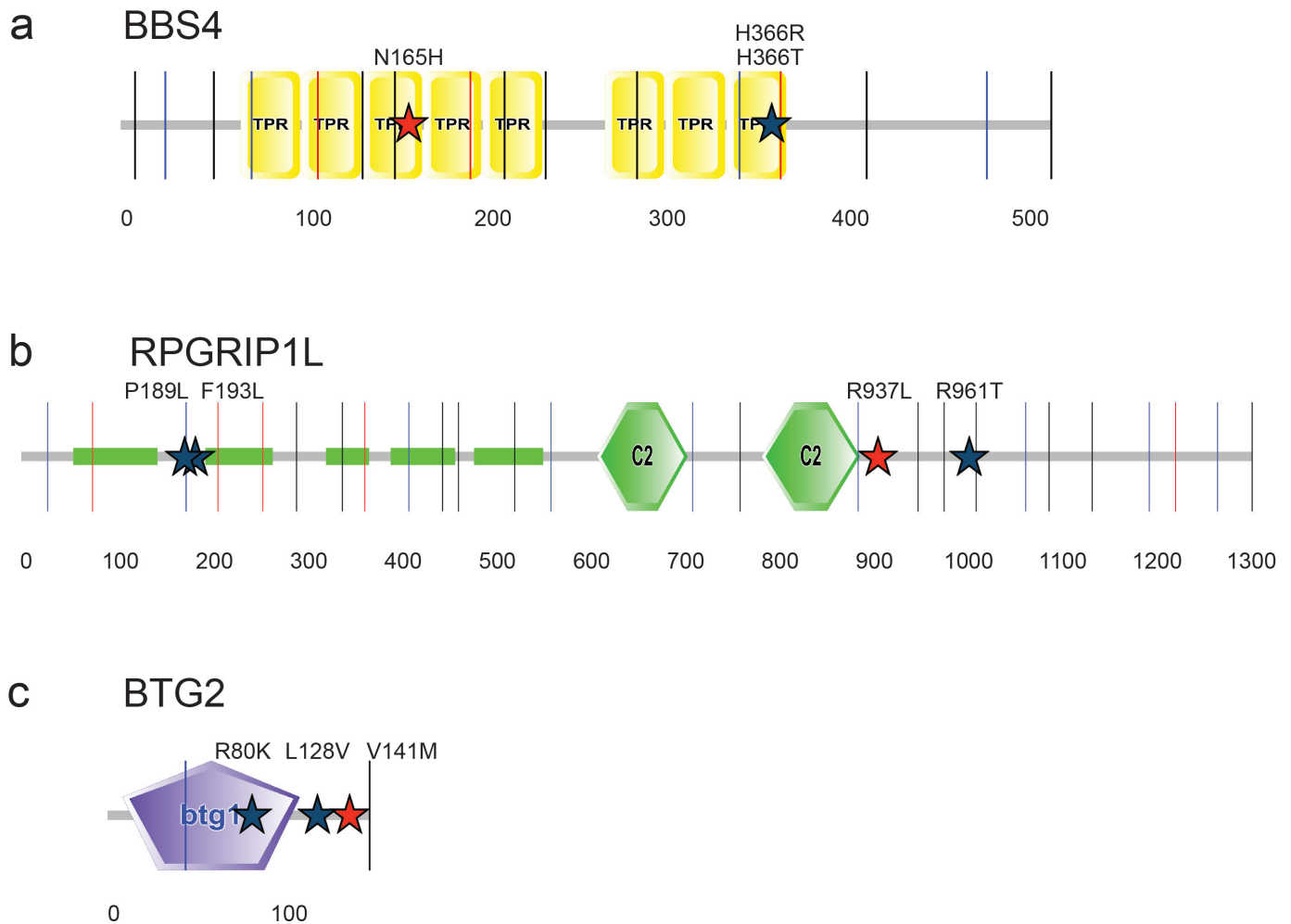
PH3-positive cell quantification was done using the Image-based Tool for Counting Nuclei (ITCN) plugin for the ImageJ software. Rolling background subtraction (25-pixel radius) and outlier removal (2.5-pixel radius, threshold = 5) were used to process images. Linear measurement of a typical cell was used to determine cell radius for ITCN analysis. Threshold level was set to 0.5. Statistical comparisons between groups were performed with a Student's *t*-test.

1. Tennesen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
2. Karolchik, D. *et al.* The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.* **42**, D764–D770 (2014).
3. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42**, D749–D755 (2014).
4. Bainbridge, M. N. *et al.* Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biol.* **12**, R68 (2011).
5. Challis, D. *et al.* An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics* **13**, 8 (2012).
6. Niederriter, A. R. *et al.* *In vivo* modeling of the morbid human genome using *Danio rerio*. *J. Vis. Exp.* **78**, e50338 (2012).



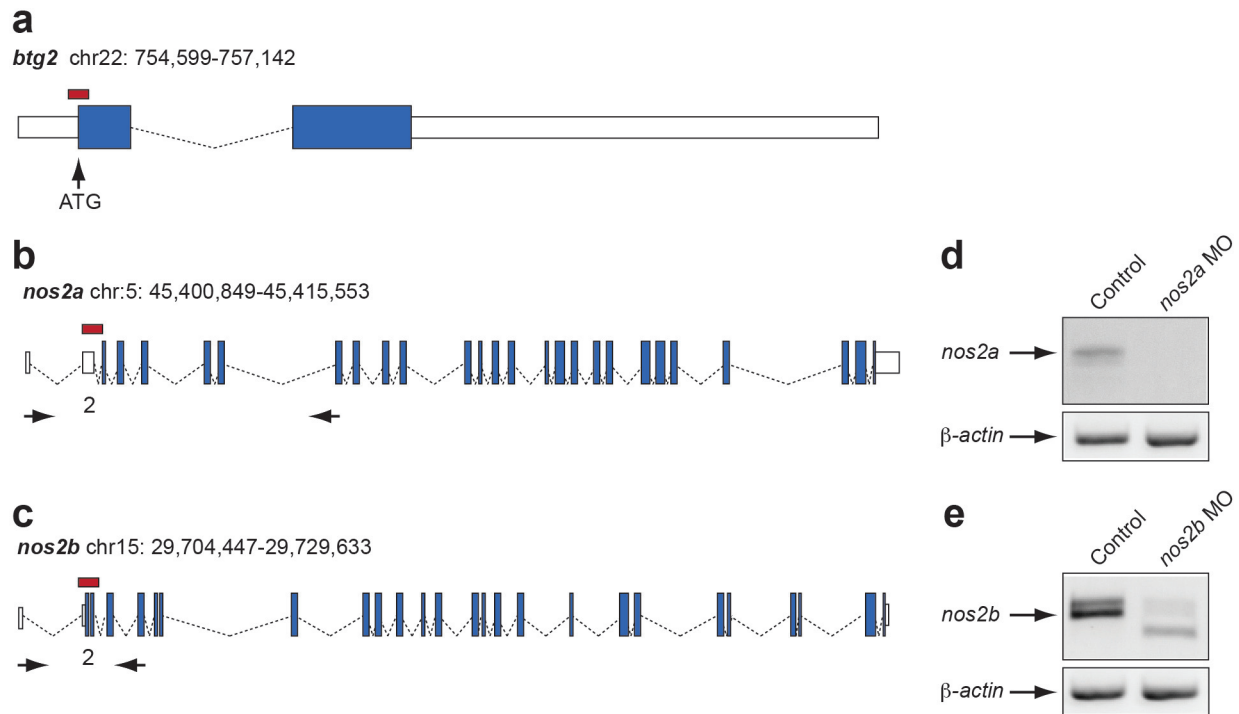
Extended Data Figure 1 | Different alignment methodologies with HumVar and ClinVar produce qualitatively similar alignments. a, b, Distributions of missense variants annotated as neutral (a) or pathogenic (b) in the HumVar and ClinVar data sets, with each of the five alignment strategies described

in the text (MultiZ unfiltered, MultiZ mammals-only, EPO, MultiZ with alignment quality filter, MultiZ with >1 sequence filter). All distributions are quantitatively similar. Compare with Fig. 2c, d.



Extended Data Figure 2 | Protein domain structure of functionally tested human disease genes. **a**, Schematic of BBS4 (519 amino acids) is depicted with eight tetratricopeptide (TPR) domains (yellow); **b**, RPGRIP1L (1,315 amino acids) has multiple coiled-coil domains (green rectangles) and two protein kinase C conserved region 2 (C2) domains (green hexagons); and **c**, BTG2

(158 amino acids) has one BTG1 domain (purple pentagon). Disease-causing alleles are shown with red stars; complementing alleles are represented with blue stars; amino acid number scale in increments of 100 is shown below each schematic.

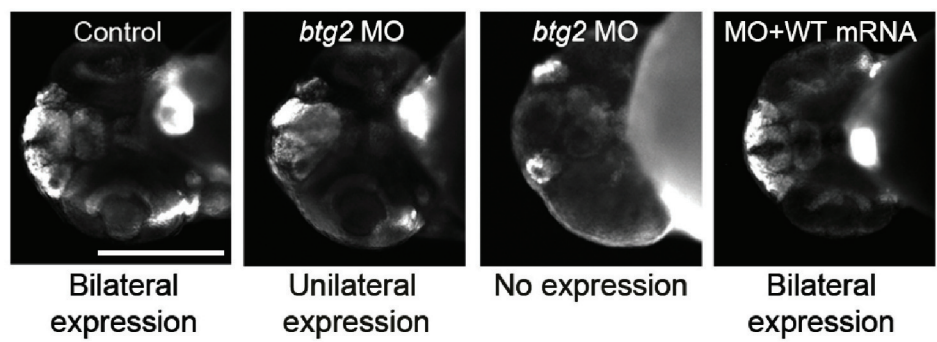


Extended Data Figure 3 | Evaluation of *btg2* and *nos2a/b* MOs.

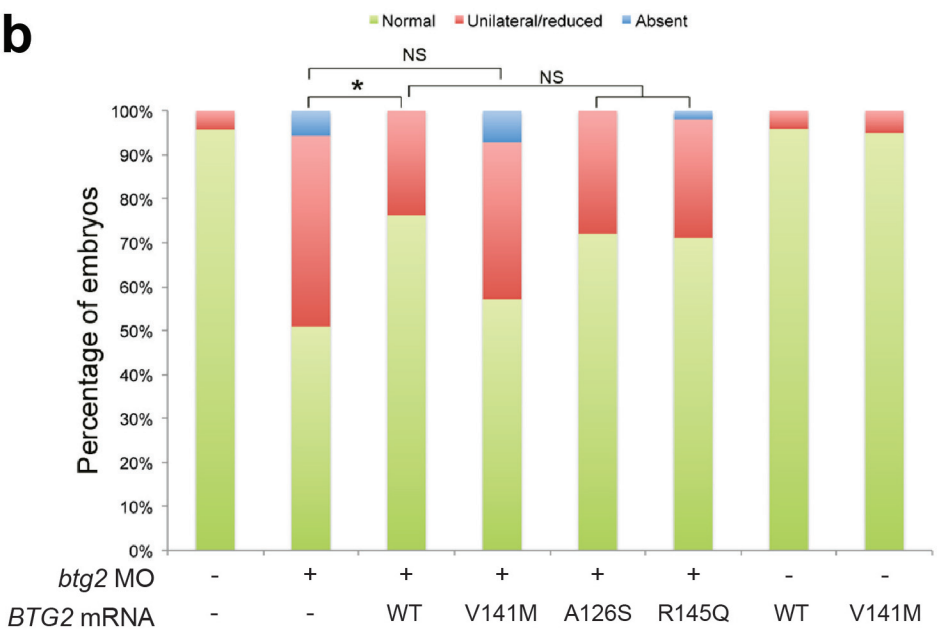
a–c, Schematic of the *D. rerio* *btg2*, *nos2a* and *nos2b* loci. Blue boxes, exons; dashed lines, introns; white boxes, untranslated regions; red boxes, MOs; ATG

indicates the translational start site; arrows, polymerase chain reaction with reverse transcription (RT-PCR) primers; number indicates the targeted exon. **d, e**, Agarose gel images of *nos2a/b* RT-PCR products.

a



b



Extended Data Figure 4 | HuC/HuD staining and quantification of 2 dpf zebrafish embryos confirms pathogenicity of BTG2 V141M. **a**, Suppression of *btg2* leads to a decrease of HuC/HuD levels at 2 dpf. Representative ventral images of control, *btg2* morphants (images show unilateral or absent HuC/HuD expression), and a rescued embryo injected with a *btg2* MO plus human BTG2 wild-type (WT) mRNA. Scale bar, 250 μ m. **b**, Percentage of embryos

with normal, bilateral HuC/HuD protein levels in the anterior forebrain or decreased/unilateral HuC/HuD protein levels in embryos injected with *btg2* MOs alone or MOs plus human BTG2 wild-type or variant mRNAs (p.V141M, index case; p.A126S and p.R145Q, control alleles). * $P < 0.05$ (two-tailed *t*-test comparisons between MO-injected and rescued embryos; $n = 38$ –78 per injection batch).

Genetic compensation induced by deleterious mutations but not gene knockdowns

Andrea Rossi^{1*}, Zacharias Kontarakis^{1*}, Claudia Gerri¹, Hendrik Nolte^{1†}, Soraya Hölper¹, Marcus Krüger^{1†} & Didier Y. R. Stainier¹

Cells sense their environment and adapt to it by fine-tuning their transcriptome. Wired into this network of gene expression control are mechanisms to compensate for gene dosage. The increasing use of reverse genetics in zebrafish, and other model systems, has revealed profound differences between the phenotypes caused by genetic mutations and those caused by gene knockdowns at many loci^{1–3}, an observation previously reported in mouse and *Arabidopsis*^{4–7}. To identify the reasons underlying the phenotypic differences between mutants and knockdowns, we generated mutations in zebrafish *egfl7*, an endothelial extracellular matrix gene of therapeutic interest, as well as in *vegfaa*. Here we show that *egfl7* mutants do not show any obvious phenotypes while animals injected with *egfl7* morpholino (morphants) exhibit severe vascular defects. We further observe that *egfl7* mutants are less sensitive than their wild-type siblings to *Egfl7* knockdown, arguing against residual protein function in the mutants or significant off-target effects of the morpholinos when used at a moderate dose. Comparing *egfl7* mutant and morphant proteomes and transcriptomes, we identify a set of proteins and genes that are upregulated in mutants but not in morphants. Among them are extracellular matrix genes that can rescue *egfl7* morphants, indicating that they could be compensating for the loss of *Egfl7* function in the phenotypically wild-type *egfl7* mutants. Moreover, *egfl7* CRISPR interference, which obstructs transcript elongation and causes severe vascular defects, does not cause the upregulation of these genes. Similarly, *vegfaa* mutants but not morphants show an upregulation of *vegfab*. Taken together, these data reveal the activation of a compensatory network to buffer against deleterious mutations, which was not observed after translational or transcriptional knockdown.

Interfering with a gene's function is a widely used strategy to decipher its role. Several different approaches have been developed over the years to achieve this goal. Yet, despite having the same goal of functional inactivation, different strategies, namely knockdown (via antisense) and knockout (via genetic inactivation), often lead to different phenotypes. These discrepancies could be caused by off-target effects of the knockdown reagents, the generation and use of hypomorphic alleles, or other and more fundamental reasons. The recent development of new genome engineering techniques, such as TAL effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPRs), is allowing the facile generation of mutations and has revived concerns over the lack of specificity of knockdown reagents^{1–3}. In several cases, toxicity due to off-target effects, induction of *p53* (also known as *tp53*) transcription, interferon response, engagement of toll-like receptors and/or saturation of the RNA interference machinery can lead to phenotypes unrelated to the silencing of the target gene^{8,9}. To investigate further whether toxicity effects are the main reason for the differences between genetic mutation and gene knockdown phenotypes, we analysed the *EGF-like-domain, multiple 7 (egfl7)* gene. The *egfl7* gene is a good candidate to address this

question because of the lack of obvious phenotypes in the mouse mutants^{10,11} and the severe vascular tube formation defects observed in knockdown experiments in zebrafish, frogs and human cells^{12–14}.

We first generated *egfl7* mutants using TALENs¹⁵ targeting exon 3, which encodes part of the EMI domain (Fig. 1a and Extended Data Fig. 1). This domain precedes other domains critical for *Egfl7* activity, including EGF domains and the leucine–valine-rich carboxy (C) terminus (Fig. 1a)¹⁶. We identified several deletion alleles including a $\Delta 3$ and a $\Delta 4$ (Fig. 1b). The *egfl7* $\Delta 3$ (hereafter *egfl7*^{s980}) allele encodes a protein that lacks a non-conserved proline at position 50 (p.P50del) while the *egfl7* $\Delta 4$ allele (hereafter *egfl7*^{s981}) is predicted to encode a truncated polypeptide containing a stretch of 29 incorrect amino acids starting with a Gln to Leu substitution at position 49 (p.Gln49Leufs*30) (Fig. 1b). To investigate the severity of these mutant alleles, we first examined *egfl7* transcript levels by quantitative PCR (qPCR). The premature stop codon in *egfl7*^{s981} led to a decrease of approximately 50% in transcript levels compared with wild-type (WT) and *egfl7*^{s980} mutant embryos, indicating an increased messenger RNA (mRNA) degradation rate (Extended Data Fig. 2a). To characterize the different *egfl7* mutant alleles further, we cloned the *egfl7* WT, s980 and s981 complementary DNAs (cDNAs) in a mammalian expression vector and transfected HUVEC cells. Unlike *Egfl7* WT and *Egfl7*^{s980}, the *Egfl7*^{s981} protein was mostly absent in the medium or the cells, suggesting that this truncated polypeptide is rapidly degraded and/or poorly translated and secreted (Extended Data Fig. 2b). Altogether, these data indicate that *egfl7*^{s981} is a severe mutant allele, possibly even a null.

To analyse *Egfl7* function during vascular development, the *egfl7*^{s981} mutant fish were crossed into the *Tg(kdrl:HRAS-mCherry)*¹⁷ and *Tg(kdrl:GFP)*¹⁸ backgrounds. We also developed a robust method based on high-resolution melt analysis to identify the different genotypes (Extended Data Fig. 1b). Surprisingly, no differences in gross morphology were evident between *egfl7* WT and mutant animals. However, a sporadic onset of brain haemorrhage was observed in fewer than 5% of the mutant animals at 72 hours post-fertilization (hpf) (Fig. 1c, d). Besides the haemorrhagic foci, no obvious abnormalities were detected in vasculogenesis, angiogenesis or circulation in any region of the brain or the rest of the body (Fig. 1e, f and Extended Data Fig. 3). Moreover, *egfl7*^{s981} mutant animals survive to become fertile adults. In summary, while *egfl7* morphants exhibit severe vascular defects¹², *egfl7* mutants exhibit very mild, if any, phenotypes.

This discrepancy between mutant and morphant phenotypes could be explained by several reasons including morpholino (MO) off-target effects. We thus sought to assess the specificity and toxicity of *egfl7* MO. First, to evaluate the effectiveness of the *egfl7* MO, we engineered the *egfl7* locus through the co-injection of TALENs and a single-stranded DNA (ssDNA) donor encoding a Myc-tag (Extended Data Fig. 4), and generated a stable transgenic line. We then injected *Egfl7* Myc-tag embryos with 1 ng of *egfl7* MO and analysed protein levels by western blot at 24 hpf. The relative expression of *Egfl7* Myc-tag was reduced by approximately 80% in the morphants compared with

¹Max Planck Institute for Heart and Lung Research, 61231 Bad Nauheim, Germany. [†]Present address: Institute for Genetics and CECAD, University of Cologne, 50931 Cologne, Germany.

*These authors contributed equally to this work.

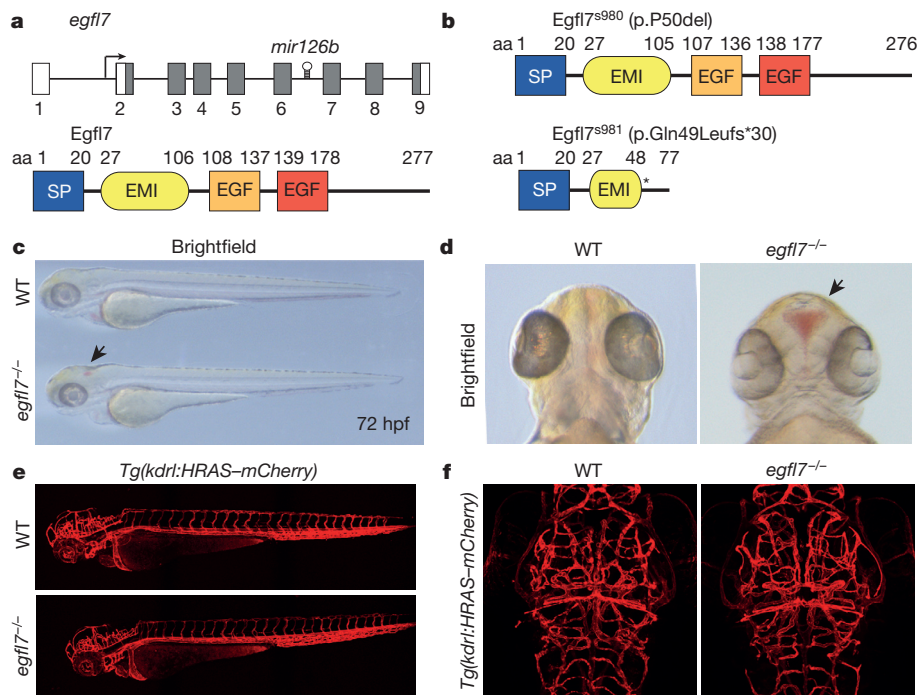


Figure 1 | Generation of zebrafish *egfl7* mutant alleles and sporadic brain haemorrhage in mutant larvae. **a**, Top: *egfl7* consists of 9 exons and *mir126b* is embedded in intron 6. The protein is encoded by exons 2–9 (grey boxes). **a**, Bottom: Egfl7, 277 amino acids (aa) long, contains a signal peptide (blue), an EMI domain (yellow), an EGF domain that contains a Delta-Serrate-LAG-2 (DSL) motif (orange) and a Ca^{2+} -binding EGF domain (red). **b**, Top: the *egfl7*^{s980} lesion ($\Delta 3$) leads to the deletion of proline at position 50 (p.P50del). **b**, Bottom: the *egfl7*^{s981} allele ($\Delta 4$) encodes a truncated 77-amino-acid-long polypeptide (p.Gln49Leufs*30) that contains a signal peptide (blue) and a partial EMI domain (yellow) followed by a frameshift leading to a premature stop codon. **c**, **d**, Brightfield micrographs of 72 hpf WT and *egfl7*^{s981/s981} larvae in lateral and ventral views. Arrows point to area of haemorrhage. **e**, **f**, Confocal micrographs of 72 hpf *Tg(kdrl:HRAS-mCherry)* WT and *egfl7*^{s981/s981} larvae in lateral and dorsal views.

uninjected embryos, revealing the ability of the MO to inhibit *egfl7* mRNA translation. A widely recognized MO off-target effect is the transcriptional activation of *p53* (ref. 9). We thus measured *p53* expression by qPCR and observed no significant difference between embryos injected with 1 ng of MO and uninjected embryos. However, *p53* expression was clearly induced in embryos injected with 2 or 4 ng of MO (Extended Data Fig. 5). We next reasoned that if the *egfl7* MO did not induce off-target effects, it should not cause defects in *egfl7* null mutants. Thus, we injected embryos obtained from *egfl7*^{s981/+} incrosses with 1 ng of *egfl7* MO. We subsequently selected and genotyped 32 embryos that showed a vascular phenotype, namely intersegmental vessel defects, reduced circulatory loop and/or pericardial oedema (Fig. 2a). Notably, we found that these embryos did not follow the Mendelian pattern observed in controls: 17 embryos were WT, 12 heterozygous and only 3 mutant (Fig. 2b, c), suggesting that the *egfl7* mutants were less sensitive than WT to MO injections. Confocal micrographs of WT, heterozygous and mutant embryos injected with 1 ng of *egfl7* MO (Fig. 2a) support this hypothesis. To investigate why some mutant embryos showed a phenotype when injected with *egfl7* MO, we repeated this experiment using a lower MO dose (0.5 ng). This experiment resulted in a clear reduction of the number of mutants in the selected population (1 mutant, 20 heterozygous and 21 WT fish out of 42 selected for vascular abnormalities, $P < 0.0001$). In the same experiment, out of ten WT-looking embryos, eight were mutant and two heterozygous for *egfl7*^{s981} ($P = 0.0003$) (data not shown), supporting the observation that the mutant fish are less sensitive to *egfl7* MO and indicating that the *egfl7* MO has minimal off-target effects at these concentrations.

To investigate the differences between the mutant and morphant phenotypes further, we used an alternative knockdown approach and took advantage of the recently developed CRISPR interference (CRISPRi) technology¹⁹ to inhibit *egfl7* transcript elongation. We designed two guide RNAs (gRNAs) to target the non-template strand of the 5' untranslated region (UTR) and exon 2 of *egfl7* as well as one gRNA targeting the template strand of exon 2 (negative control) (Extended Data Fig. 6a). The relative *egfl7* expression levels were then quantified at 20 hpf using qPCR on pools of ten embryos injected with gRNAs and catalytically inactive (dead) CAS9 (dCAS9). Non-template gRNAs were able to inhibit *egfl7* transcript levels by approximately 60% compared with uninjected or template gRNA-injected embryos

(Extended Data Fig. 6b). *Tg(kdrl:GFP)* embryos injected with gRNAs and dCAS9 exhibited different degrees of vascular abnormalities at 48 hpf, including intersegmental vessel defects, reduced circulatory loop and pericardial oedema after non-template but not template gRNA injections (Extended Data Fig. 6c). Altogether, these data show that transcriptional or translational knockdown of *egfl7* can lead to severe cardiovascular phenotypes while a severe genetic lesion does not.

To identify molecules underlying the different phenotypes observed in mutants versus morphants, we performed mass spectrometry and RNA profiling analyses in *egfl7* WT, homozygous mutant (*egfl7*^{s981}) and morphant embryos at 24 hpf. We assessed the proteomes by 4 h 'single shot' liquid chromatography–tandem mass spectrometry (LC–MS/MS) and identified more than 6,000 proteins with high reproducibility ($r > 0.90$ for biological and technical replicates between mutants and WT; Extended Data Fig. 7). To identify significant differences in individual protein expression, we used randomization-based false detection rate (FDR) estimation for multiple-testing correction and identified only one protein differentially expressed between mutants and WT (Fig. 3a). Strong upregulation (more than fivefold) was found for Emilin3a, suggesting its possible role in compensation. Additionally, we found no significant upregulation of Emilin3a in morphants compared with WT (Fig. 3b; Extended Data Fig. 8a), further highlighting Emilin3a as a possible compensating protein. Moreover, RNA sequencing (RNA-seq) and qPCR analyses indicated that not only *emilin3a* but also *emilin3b* and *emilin2a* were upregulated in mutants but not in morphants or CRISPRi injected embryos (Fig. 3c; Extended Data Fig. 8b). Interestingly, all these proteins contain an EMI domain, one of the key units of Egfl7 function¹⁶, and, like Egfl7, can regulate elastogenesis^{20,21}. We then reasoned that if Emilins are able to functionally replace Egfl7, they might rescue *egfl7* morphants. Embryos were injected with *egfl7* MO or co-injected with *egfl7* MO and *egfl7*, *egfl7*^{s981}, *Emilin2* or *Emilin3* mRNA and screened for circulatory loop defects at 48 hpf. Similarly to *egfl7* mRNA, *Emilin2* and *Emilin3* mRNAs were both able to rescue the circulatory defects in a significant proportion of *egfl7* morphants, while *egfl7*^{s981} mRNA was not (Fig. 4). These results support the hypothesis that the upregulation of *emilin* genes in *egfl7*^{s981} mutants is at least partly responsible for their lack of phenotype. To test whether the transcriptional changes we identified in mutants but not morphants were a peculiarity of the *egfl7* locus, we generated TALEN mutants for *vegfaa* (data not shown).

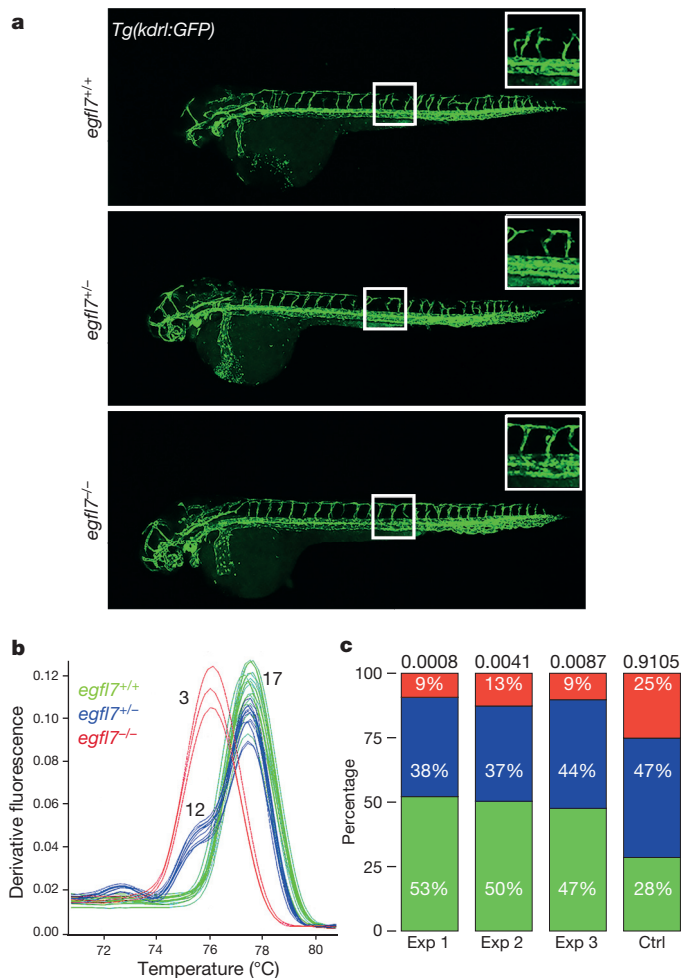


Figure 2 | Zebrafish *egfl7* mutant embryos are less sensitive to *egfl7* morpholino injections. **a**, Confocal micrographs of *Tg(kdrl:GFP)* WT, *egfl7*^{981/+} and *egfl7*^{-/-} 48 hpf embryos injected with 1 ng of *egfl7* MO (AS₄₇) in lateral views. **b**, High-resolution melt analysis genotyping example of 32 embryos (from an *egfl7*^{981/+} incross) selected for vascular defects at 48 hpf, showing the melting curves of 17 *egfl7* WT (green), 12 *egfl7*^{981/+} (blue) and 3 *egfl7*^{981/981} (red) embryos. **c**, Genotype distribution (at 48 hpf) of *egfl7*^{981/+} incross progeny injected with 1 ng of *egfl7* MO at the one-cell stage and subsequently selected for the vascular phenotypes (independent experiments (Exp 1, 2 and 3) or randomly selected (Ctrl)). The population of randomly selected embryos follows the expected Mendelian ratio, but the phenotype-selected populations show significant skewing towards *egfl7* WT. *P* value represents two-tailed value for χ^2 test with two degrees of freedom; *n* = 32 embryos genotyped in each experiment. Note that the *egfl7*^{981/+} embryos are also underrepresented in the phenotype-selected populations (corresponding *P* values for experiments 1, 2 and 3 are 0.0033, 0.0066 and 0.032, respectively).

Interestingly, qPCR analysis showed that *vegfab*, a paralogue of *vegfaa*, was upregulated in mutants but not morphants (Extended Data Fig. 9a). Additionally, blocking *Vegfaa* function using a dominant negative approach also failed to trigger *vegfab* upregulation, placing the signal triggering compensation upstream of protein function (Extended Data Fig. 9b).

Concerns have been raised over the use of antisense reagents, including MOs, as they may cause off-target effects and lead to aberrant conclusions. This debate was recently revived by the generation of mutations in many genes whose function was previously studied using MOs; strikingly, a majority of the resulting mutants exhibit a different phenotype from the one reported for the corresponding morphants; in fact, most often the mutants exhibit no obvious phenotype¹. In our study, we show that, at least for some genes, the phenotypic differences between mutants and morphants can be due to the activation of genetic

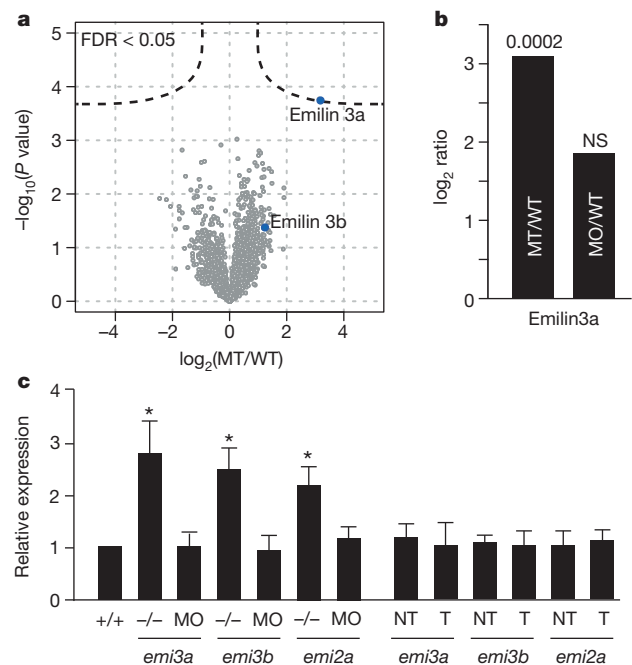


Figure 3 | Emilin3a is upregulated in mutant but not in morphant or CRISPRi embryos. **a**, Volcano plot showing significantly dysregulated proteins between 24 hpf *egfl7* WT and *egfl7*⁹⁸¹ mutant embryos using label-free quantification. Emilin3a and Emilin3b are highlighted in blue. **b**, Morphants did not show a significant upregulation of Emilin3a in unbiased mass-spectrometry-based proteomics comparing mutants, WT and morphants. A two-sided *t*-test was used to assess *P* values and FDR was controlled by a randomization-based SAM approach. **c**, mRNA expression of *emilin3a*, *emilin3b* and *emilin2a* in *egfl7* WT, mutant, morphant and CRISPRi (template and non-template strand) embryos at 20 hpf; qPCR data, pools of 20–30 embryos each, expression normalized to *gapdh* (WT expression set at 1 for each gene). The *emilin* genes were upregulated in *egfl7*⁹⁸¹ mutants but not after translational or transcriptional inhibition. **P* ≤ 0.05.

compensation in the former but not the latter. We show here that the upregulation of Emilins can compensate for the loss of Egfl7, but anticipate that other genes are involved in this process. On the basis of our data, we propose two additional recommendations for using MOs: the first is to use doses that do not induce *p53* expression as in many cases this induction indicates off-target effects; the second is to titrate the MO dose so that it does not cause additional phenotypes in a null mutant background, as such phenotypes would be due to non-specific effects. The mechanisms underlying the compensation observed in mutants but not in morphants are likely to be complex and so will their investigation. Interestingly, we observed no upregulation of the *emilin* genes in the $\Delta 13$ (*s980*) allele, suggesting that a non-deleterious genomic lesion is not sufficient to trigger this response. On the other hand, we observed *emilin* gene upregulation in embryos injected with *egfl7* TALENs, indicating that a deleterious mutation does not need to go through the germline to trigger this response. We also detected *emilin* gene upregulation in embryos carrying only one *egfl7*⁹⁸¹ mutant allele (data not shown). This observation might explain the partial protection of heterozygous embryos from *egfl7* MO injections (Fig. 2).

In summary, our data show that, for *egfl7*, one can identify a dose of MO that has no effect in most *egfl7* mutant embryos but causes clear vascular defects in WT, indicating that these morphant phenotypes are not due to off-target effects. Further, *egfl7* mutants show no phenotypes but exhibit a clear upregulation of several members of the *emilin* gene family. These Emilin proteins share an important functional domain with Egfl7, and, probably with additional proteins, can compensate for the loss of Egfl7 function. Notably, a recent study of the Icelandic population identified individuals with homozygous loss-of-function mutations in *EGFL7* (ref. 22), indicating that compensation

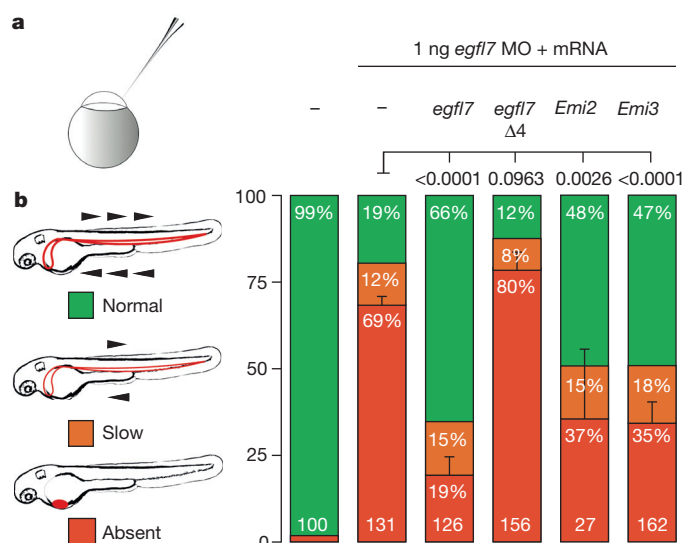


Figure 4 | *Emilin2* and *Emilin3* can rescue *egfl7* morphants. **a**, Design of the rescue experiment. One nanogram of *egfl7* MO was injected in WT embryos, alone or together with 400 pg of mRNA (*egfl7* WT, *egfl7* $\Delta 4$, *Emilin2* or *Emilin3*). **b**, Injected embryos were sorted according to their circulatory loop phenotype into three classes: normal, slow and absent circulation. Injection of *egfl7* MO resulted in 69% of embryos lacking circulation at 48 hpf. This percentage was reduced to 19% when co-injecting *egfl7* mRNA, and to 37% and 35% when co-injecting *Emilin2* and *Emilin3* mRNA, respectively. In contrast, mRNA from the *egfl7* $\Delta 4$ mutant allele did not rescue (80% of embryos lacked circulation). Uninjected siblings are shown for comparison (99% normal). Number at the bottom of each bar is the total number of embryos from two independent experiments. Error bars, s.e.m. for the 'absent circulation' class. P value represents two-tailed value for Fisher's exact test.

for severe lesions at this locus might also be at play in humans. It will be interesting to determine whether the upregulation of *EMILIN* genes is also present in these individuals. Of course, detailed studies will be needed to determine whether such compensation is the reason for the phenotypic differences between mutants and morphants for other genes. More importantly, our study illustrates the power of comparing mutants and morphants to identify modifier genes, a goal that remains a major challenge in the field of human genetics.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 December 2014; accepted 22 May 2015.

Published online 13 July 2015.

- Kok, F. O. *et al.* Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev. Cell* **32**, 97–108 (2015).

- Stainier, D. Y., Kontarakis, Z. & Rossi, A. Making sense of anti-sense data. *Dev. Cell* **32**, 7–8 (2015).
- Law, S. H. & Sargent, T. D. The serine-threonine protein kinase PAK4 is dispensable in zebrafish: identification of a morpholino-generated pseudophenotype. *PLoS ONE* **9**, e100268 (2014).
- Daude, N. *et al.* Knockout of the prion protein (PrP)-like *Sprrn* gene does not produce embryonic lethality in combination with PrP(C)-deficiency. *Proc. Natl Acad. Sci. USA* **109**, 9035–9040 (2012).
- De Souza, A. T. *et al.* Transcriptional and phenotypic comparisons of *Ppara* knockout and siRNA knockdown mice. *Nucleic Acids Res.* **34**, 4486–4494 (2006).
- Smart, N. & Riley, P. R. Thymosin $\beta 4$ in vascular development, response to research commentary. *Circ. Res.* **112**, 29–30 (2013).
- Gao, Y. *et al.* Auxin binding protein 1 (ABP1) is not required for either auxin signaling or *Arabidopsis* development. *Proc. Natl Acad. Sci. USA* **112**, 2275–2280 (2015).
- Anon. RNA interference on target. *Nature Methods* **3**, 659 (2006).
- Robu, M. E. *et al.* p53 activation by knockdown technologies. *PLoS Genet.* **3**, e78 (2007).
- Schmidt, M. *et al.* EGFL7 regulates the collective migration of endothelial cells by restricting their spatial distribution. *Development* **134**, 2913–2923 (2007).
- Kuhnert, F. *et al.* Attribution of vascular phenotypes of the murine *Egfl7* locus to the microRNA *miR-126*. *Development* **135**, 3989–3993 (2008).
- Parker, L. S. *et al.* The endothelial-cell-derived secreted factor Egfl7 regulates vascular tube formation. *Nature* **428**, 754–758 (2004).
- Charpentier, M. S. *et al.* CASZ1 promotes vascular assembly and morphogenesis through the direct regulation of an EGFL7/RhoA-mediated pathway. *Dev. Cell* **25**, 132–143 (2013).
- Huang, C. *et al.* VE-statin/Egfl7 siRNA inhibits angiogenesis in malignant glioma *in vitro*. *Int. J. Clin. Exp. Pathol.* **7**, 1077–1084 (2014).
- Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39**, e82 (2011).
- Nichol, D. & Stuhlmann, H. EGFL7: a unique angiogenic signaling factor in vascular development and disease. *Blood* **119**, 1345–1352 (2012).
- Chi, N. C. *et al.* Foxn4 directly regulates *tbx2b* expression and atrioventricular canal formation. *Genes Dev.* **22**, 734–739 (2008).
- Jin, S. W., Beis, D., Mitchell, T., Chen, J. N. & Stainier, D. Y. Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development* **132**, 5199–5209 (2005).
- Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols* **8**, 2180–2196 (2013).
- Zanetti, M. *et al.* EMILIN-1 deficiency induces elastogenesis and vascular cell defects. *Mol. Cell. Biol.* **24**, 638–650 (2004).
- Leievre, E. *et al.* VE-statin/egfl7 regulates vascular elastogenesis by interacting with lysyl oxidases. *EMBO J.* **27**, 1658–1670 (2008).
- Sulem, P. *et al.* Identification of a large set of rare complete human knockouts. *Nature Genet.* **47**, 448–452 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H.-B. Kwon and other members of the laboratory, past and present, as well as K. Sampath, D. Wainstock, C. Moens and M. Grether, for discussions, comments on the manuscript and/or reagents, and the Max Planck Society, Packard foundation, and EMBO for funding.

Author Contributions All authors were involved in the experimental design, data analysis and writing. Experiments were performed by all except M.K. and D.Y.R.S., who also supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.Y.R.S. (didier.stainier@mpi-bn.mpg.de).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment, except for the data shown in Fig. 2 where the inherent design of the experiment includes a blinding component.

Zebrafish handling. All zebrafish husbandry was performed under standard conditions in accordance with institutional (UCSF and MPG) and national ethical and animal welfare guidelines.

Confocal microscopy. An LSM 700 confocal laser scanning microscope (Zeiss) was used for live imaging. Embryos and larvae were anaesthetized with a low dose of tricaine, placed in a glass-bottomed Petri dish (MatTek) with a layer of 1.2% low melt agarose and imaged using Plan-Apochromat 10×/0.45 and LCI Plan-Neofluar 25×/0.8 objective lenses. Vessel integrity and permeability were analysed using micro-angiography. Fluorescein isothiocyanate (FITC)-dextran, 2,000 kDa (Sigma) was injected into the posterior cardinal vein at 48 or 72 hpf and imaged after 10 min.

Plasmids. Total RNA extraction was performed using TRIzol (Life Technologies) and used for cDNA synthesis using SuperScript second strand (Life Technologies). cDNAs encoding the Eglf7 WT, Eglf7⁹⁸⁰ and Eglf7⁹⁸¹ proteins were PCR-amplified using whole embryo cDNA as template. PCR fragments were ligated into the mammalian expression vector pcDNA3.1 myc-HIS tag between EcoRI and XhoI. All constructs were verified by sequencing. pCMV-6 plasmids containing mouse elastin microfibril interfacer 2 (*Emilin2*) or mouse elastin microfibril interfacer 3 (*Emilin3*) cDNAs were purchased from Origene. pcDNA3.1 and pCMV-6 plasmids were respectively linearized using SmaI and AgeI and *in vitro* transcribed using the mMESSAGE mMACHINE T7 kit (Life Technologies).

Cell culture and transfection. Authenticated HEK293FT (human embryonic kidney) (Life Technologies, R700-07) and HUVEC (human umbilical vein endothelial cells) cells were cultured at 37 °C in 5% CO₂, 95% air in appropriate medium containing 10% fetal bovine serum, 100 units per millilitre penicillin and 100 µg ml⁻¹ streptomycin. HEK293FT cells were used for biochemical studies because they are easy to grow and transfect, and have been used widely for cell biology research. All cell lines are routinely tested for mycoplasma in our facilities, and only mycoplasma-free cells were used in this study. Cells were transfected with cDNAs in antibiotic-free medium 12–24 h before protein extraction, using FuGene HD (Roche) at a 3:1 ratio (µl:µg nucleic acid) and 0.18 µg DNA per square centimetre. Cells were lysed in RIPA buffer and extracellular proteins precipitated with TCA (final 20%). Samples were resuspended in Laemmli buffer before gel electrophoresis.

Genome editing. TALENs were designed targeting *egfl7* (Extended Data Fig. 1) using TALEN targeter²³ (<https://tale-nt.cac.cornell.edu/>) and constructed using Golden Gate assembly¹⁵. Zebrafish embryos were injected into the cell at the one-cell stage with 100 pg total TALEN RNA.

Genome engineering was performed as previously described²⁴. The C terminus *egfl7* TALEN recognition sites are TGCTGGTAGACATCATC and TTGCAGTA GTGACTAGT. Between the binding sites is a 17-bp spacer with the 'tag' stop codon underlined (aggaaaactagacgac). A ssDNA oligonucleotide (Supplementary Table 1) was designed to target the spacer sequence between the cutting sites. A Myc-tag sequence, flanked by XhoI restriction sites, was introduced in the centre of the oligonucleotide resulting in 25-base homology arms on the 5' and 3' ends. Polyacrylamide gel electrophoresis (PAGE)-purified oligonucleotides were obtained from Sigma. One-cell-stage embryos were injected with 100–200 pg TALEN mRNA and 75 pg ssDNA donor. Screening for founders was conducted using PCR followed by XhoI restriction enzyme digest and subsequently by sequencing.

CRISPR interference. gRNA and CAS9 plasmids²⁵ were purchased from Addgene. Dead Cas9 was generated using the zebrafish-codon-optimized WT CAS (pT3TS-nls-zCas9-nls; nls, nuclear localization signal) as a template. The D10A and H840A mutations were generated using the primers in Supplementary Table 1. Site-directed mutagenesis was performed using PfuUltra Fusion HS (Agilent). The PCR reaction protocol was 95 °C for 1 min, then 18 cycles of 95 °C for 50 s, 60 °C for 50 s and 68 °C for 1 min per kilobase of plasmid length, then 68 °C for 7 min and 4 °C hold. Dpn1 (1 µl) was added to the PCR reaction and incubated at 37 °C for 1 h to digest parental DNA and transform into competent cells. CRISPR gRNAs were designed using CRISPR design (<http://crispr.mit.edu/>) (Zhang laboratory). Oligonucleotides were annealed in a thermo block at 90–95 °C for 5 min followed by a slow cooling to room temperature (~20 °C) and cloned in gRNA plasmid between BsmBI sites. All constructs were verified by sequencing. To make nls-zCas9-nls RNA, the template DNA (pT3TS-nls-zCas9-nls) was linearized by XbaI digestion and purified using a QIAprep column (Qiagen). Capped nls-zCas9-nls RNA was synthesized using a mMESSAGE

mMACHINE T3 kit (Life Technologies) and purified using an RNA Clean and Concentrator kit (Zymo Research). To make gRNA, the template DNA was linearized by BamHI digestion and purified using a QIAprep column. gRNA was generated by *in vitro* transcription using a T7 RNA polymerase MEGA short script T7 kit (Life Technologies). After *in vitro* transcription, the gRNA (approx 100 nucleotides long) was purified using RNA clean and concentrator (Zymo Research). dCAS9 mRNA (100–400 pg) and gRNA (50–100 pg) were co-injected in the cell at the one-cell stage and at least ten pooled embryos were used to evaluate the expression level of the targeted genes by qPCR. Initial experiments were performed with gRNAs targeting the *tnt2a* gene. In general, a substantial increase in knockdown efficiency was observed when combining multiple guides, indicating a synergistic effect. The *egfl7* gene was knocked down by using two to four gRNAs (Supplementary Table 1).

Microinjection of morpholinos. The ATG morpholinos *egfl7* (5'-CAGGTGT GTCTGACAGCAGAAAGAG-3'), *vegfaa* (5'-GTATCAAATAAACACCAA GTTCAT-3') and *tp53* (5'-GCGCCATTGCTTTGCAAGAAATTG-3'), were purchased from GeneTools and injected at the indicated amounts (0.5, 1, 2 or 4 ng for *egfl7*, 2 ng for *vegfaa* and 1 ng for *tp53*). To identify the potential effects of *p53* induction in *egfl7* morphants, we compared the phenotype of embryos co-injected with *egfl7* and *p53* MO or *egfl7* MO alone (1 ng for each MO), and did not detect any obvious differences. The experiments testing the *egfl7* MO effect on *egfl7* mutants were blind (injection into fertilized eggs from an incross of heterozygous fish followed by phenotyping and then genotyping). The *egfl7* morphant rescue experiments were not blind. Sample sizes for these and other experiments were determined on the basis of previous experience.

Genotyping. Embryos or fin-clips were placed in PCR tubes, with 50 µl of elution buffer (10 mM Tris-Cl, pH 8.5) and 1 mg ml⁻¹ proteinase K added to each well and then incubated at 55 °C for 2 h. The samples were then heated to 95 °C for 10 min to inactivate proteinase K. Primers were designed using primer3: http://biotools.umassmed.edu/bioapps/primer3_www.cgi. An Eco Real-Time PCR System (Illumina) was used for the PCR reactions and high-resolution melt analysis. DyNamo SYBR green (Thermo Fisher Scientific) was used in these experiments. PCR reaction protocols were 95 °C for 15 s, then 40 cycles of 95 °C for 2 s, 60 °C for 2 s and 72 °C for 2 s. Following the PCR, a high-resolution melt curve was generated by collecting SYBR-green fluorescence data in the 65–95 °C range. The analyses were performed on normalized derivative plots.

Electrophoresis. Laemmli SDS-PAGE gels consisted of a 4–20% running gel and 3% stacking gel or Tricine SDS-PAGE 12%. For immunoblots, membranes were blocked with 5% non-fat milk and incubated at 4 °C overnight with mouse (SC-40) or rabbit (SC-789) anti-Myc antibody (Santa Cruz Biotechnology), or anti- α -tubulin (T9026, Sigma). Membranes were then rinsed, incubated for 1 h with horseradish-peroxidase-conjugated anti-rabbit immunoglobulin-G (IgG) or anti-mouse IgG (Santa Cruz Biotechnology), rinsed extensively, and labelled proteins were detected using the Clarity Western substrate (Biorad).

Mass spectrometry. Embryos (*egfl7* WT, *egfl7*⁹⁸¹ mutants and morphants) at 20–24 hpf were lysed in 6 M urea and 2 M thiourea (in HEPES buffer pH = 8.5). The lysate was clarified by centrifugation and proteins were subjected to in-solution digestion. In brief, proteins were reduced (0.1 M DTT, 30 min at room temperature) and alkylated (55 mM IAA, 30 min at room temperature in the dark). Lys-C was added in a 1:100 enzyme:protein ratio and incubated for 3 h. Urea concentration was diluted to 2 M using 50 mM ammonium bicarbonate, and trypsin (Promega) was added in a 1:100 enzyme:protein ratio. After incubation for 18 h, generated peptides were de-salted using the Stop and Go Extraction tip technology before mass spectrometric analysis. All WT and *egfl7*⁹⁸¹ mutant experiments were performed at least in technical and biological duplicates. For morphant embryos, we measured protein changes in technical triplicates after pooling more than 20 embryos, thus reducing biological variability. The instrumentation for LC-MS/MS analysis consisted of a nano LC 1000 (Proxeon, now Thermo Scientific) coupled via a nano-electrospray ionization source to a quadrupole-based bench-top QExactive Plus or QExactive mass spectrometer. Separation of peptides according to their hydrophobicities was achieved on a 50 cm in-house packed column (internal diameter 75 µm, C18 Beads (Dr. Maisch) diameter 1.8 µm) using a binary buffer system: (A) 0.1% formic acid in H₂O and (B) 0.1% formic acid in 80% acetonitrile. A linear gradient within 220 min from 8% to 38% of B, followed by an exponential increase to 90% B and a re-equilibration step to 5% B within 20 min, was used for peptide elution. Mass spectra were acquired at a resolution of 70,000 (200 *m/z*) using an AGC target of 1E6 and a maximal injection time of 20 ms. The ion-ten method was applied for subsequent acquisition of high-energy collision-induced dissociation (HCD) fragmentation MS/MS spectra of the ten most intense peaks. Resolution was set to 17,500 at 200 *m/z* and 5E5 ions (AGC target) were collected in the C-trap within a maximal injection time of 60 ms using an isolation window of 1.3 thomsons (Th) (1 Th = 1.036426 × 10⁻⁸ kg C⁻¹)²⁶. Raw files were processed using MaxQuant 1.4.1.2 (ref. 27) and the implemented

Andromeda search engine²⁸. For peptide assignment, MS/MS HCD fragmentation spectra were correlated to the Uniprot *Danio rerio* database (2014). A list of common contaminants was included in the searches that were performed with tryptic specificity. Default settings were used for MS and MS/MS mass tolerances and peptide length. The FDR was set to 1% on protein and peptide levels and estimated by the implemented decoy algorithm. Oxidation of methionine residues and acetylation on the protein N-term were set as variable modifications, and carbamidomethyl at cysteine residues was defined as a fixed modification. The match-between-runs, label-free quantification and re-quantification were enabled. Statistical analysis and data visualization were done in the environment R.

The package *siggenes* from Bioconductor was used to determine significance of proteome changes at a FDR cutoff of less than 0.05 (ref. 29). Note that the protein list was filtered for at least 50% quantification over all experiments. In label-free protein quantification, a common problem is that low abundance proteins are likely to be not quantifiable, leading to a right-shifted normal distribution. Thus, missing values were replaced along a Gaussian distribution using a log₂ downshift of 1 and a width of 0.4. Imputation was inspected by histograms to mimic a Gaussian distribution for the complete data set (columnwise) to avoid too high a frequency of low-intensity values. Significance was assessed by a two-sided *t*-test of log₂ intensity values. Note that to compare morphants with WT, we used technical replicates of pooled embryos against all experiments for WT embryos. Five hundred randomizations were used to estimate FDR, using a cutoff of 0.05 while *s0* (the fudge factor) was defined as 0.1. Protein ratios were calculated by subtracting the average of the respective groups. Data are shown in Supplementary Table 2.

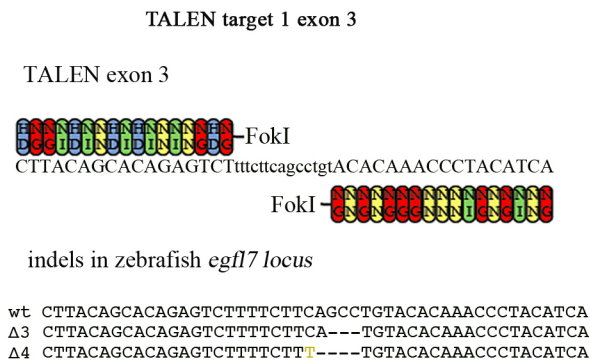
RNA profiling. Total RNA from *egfl7* WT, *egfl7⁹⁸¹* mutants and morphants at 24 hpf was prepared using TRIzol (Life Technologies). RNA profiling was performed by ZF-screens using an Illumina HiSeq 2500 ultra-high-throughput sequencing system. Data are shown in Supplementary Table 3.

qPCR. An Eco Real-Time PCR System (Illumina) was used for qPCR experiments. Gene expression was normalized relative to *gapdh*. All reactions were performed in technical triplicates; the results represent biological triplicates (unless otherwise stated) including the s.e.m. Supplementary Table 1 shows the primers used for these experiments.

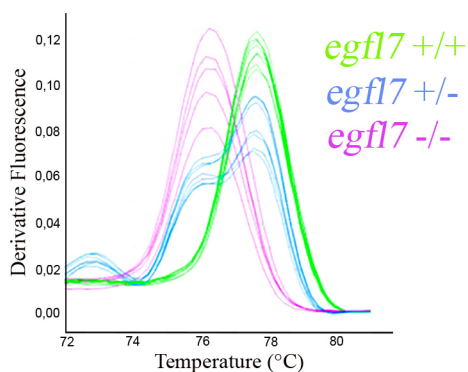
Additional data. Mature miR126 levels were quantified in WT and *egfl7^{981/981}* embryos using the miRNA QRT-PCR Detection Kit (Agilent). No significant changes were observed at 48 or 72 hpf. Maternal zygotic *egfl7⁹⁸¹* mutant embryos were generated by incrossing homozygous mutant adults; they exhibited no additional phenotypes compared with zygotic mutant embryos. In addition, we observed no evidence of maternal *egfl7* mRNA by RT-PCR. Mutant samples for proteomics, RNA-seq and qPCR analyses were MZ mutants.

23. Doyle, E. L. *et al.* TAL effector specificity for base O of the DNA target is altered in a complex, effector- and assay-dependent manner by substitutions for the tryptophan in cryptic repeat -1. *PLoS ONE* **8**, e82120 (2013).
24. Bedell, V. M. *et al.* In vivo genome editing using a high-efficiency TALEN system. *Nature* **491**, 114–118 (2012).
25. Jao, L. E., Wente, S. R. & Chen, W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl Acad. Sci. USA* **110**, 13904–13909 (2013).
26. Nolte, H. *et al.* Global protein expression profiling of zebrafish organs based on in vivo incorporation of stable isotopes. *J. Proteome Res.* **13**, 2162–2174 (2014).
27. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26**, 1367–1372 (2008).
28. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
29. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).

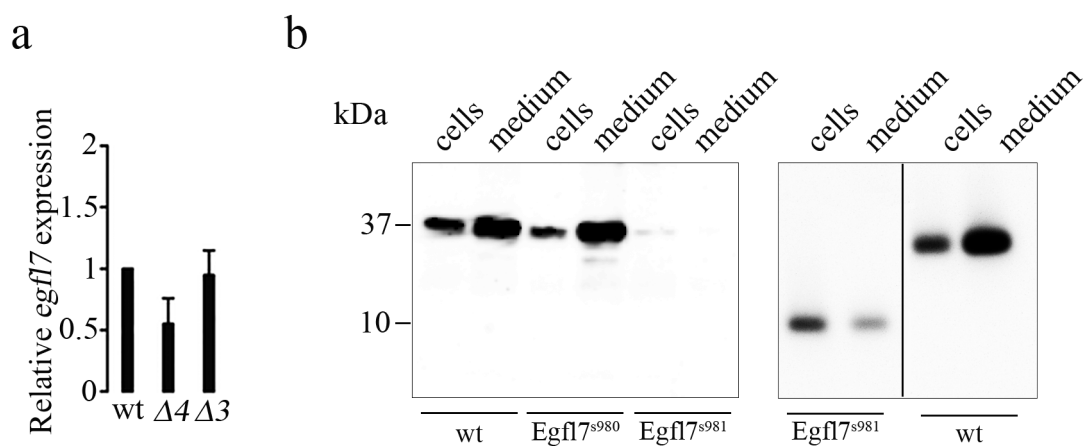
a



b

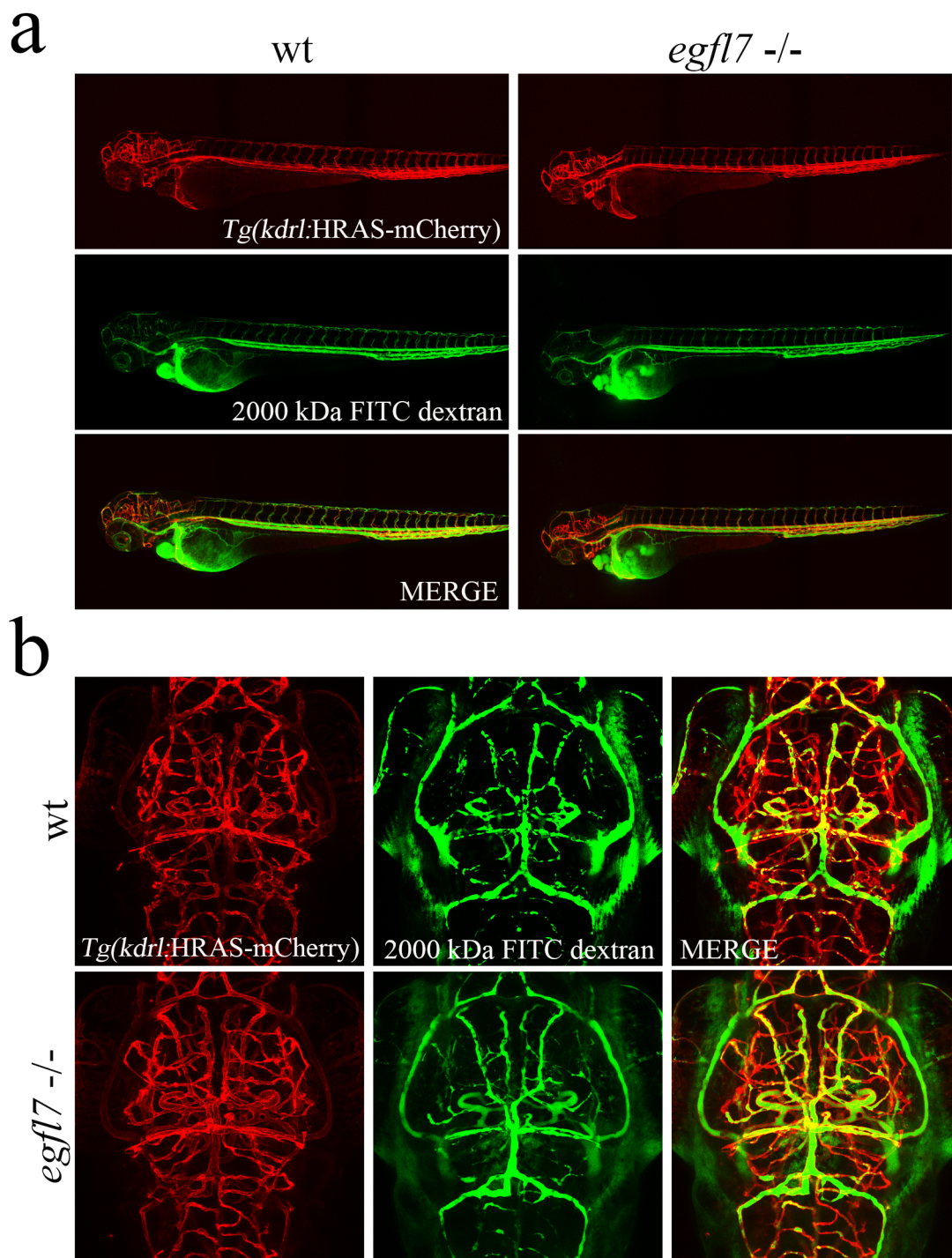


Extended Data Figure 1 | Generation and identification of zebrafish *egfl7* mutant alleles. **a**, TALENs were designed to target exon 3 of *egfl7* which encodes part of the EMI domain. Sequence alignment of part of exon 3 from the WT, *egfl7*^{s980} and *egfl7*^{s981} alleles shows TALEN indels: Δ3/s980 (three nucleotide deletion) and Δ4/s981 (five nucleotide deletion), and one nucleotide insertion (yellow). **b**, Genotyping example of single embryos sampled from a population of *egfl7* WT, *egfl7*^{s981/+} and *egfl7*^{s981/s981} fish using high-resolution melt analysis. The green curve corresponds to the WT allele and the red one to the *egfl7*^{s981} allele. Heterozygous embryos have both alleles and thus the melting profile (in blue) is a composition of the WT and mutant curves.



Extended Data Figure 2 | The *egfl7*^{s981} mutation leads to *egfl7* mRNA degradation, reduced protein expression and impaired secretion. **a**, The *egfl7*^{s981} mutation leads to *egfl7* mRNA degradation: *egfl7* mRNA expression in 24 hpf WT, *egfl7*^{s981/s981} and *egfl7*^{s980/s980} embryos. Expression normalized to *gapdh*. **b**, The *egfl7*^{s981} (p.Gln49Leufs*30) mutation leads to strongly reduced protein expression. Western blot analyses of Egfl7-Myc-tag expression in

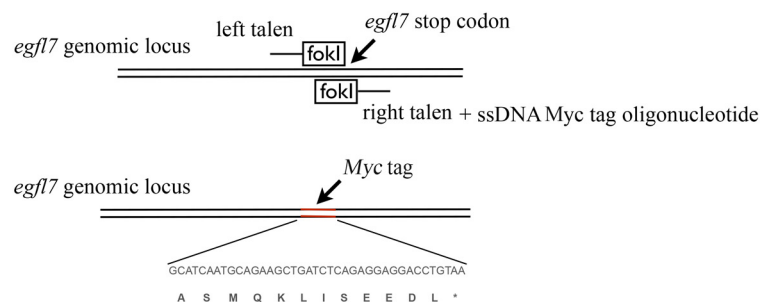
transfected HUVEC cells. Egfl7 WT and Egfl7^{s980} protein expression was strongly detected in the medium whereas the Egfl7^{s981} isoform was strongly reduced in the cells and very poorly secreted (right), or undetectable in both (left). Furthermore, Egfl7^{s981} shares high similarity to the truncated protein produced in the original *Egfl7* mutant mouse in which the protein was not detectable using an Egfl7 antibody¹⁰.



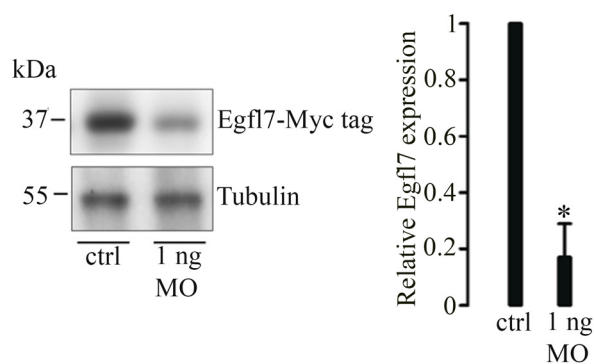
Extended Data Figure 3 | Vessel integrity and permeability do not appear to be affected in *egfl7*^{981/s981} larvae. A fluorescent molecule (2000 kDa FITC-dextran) was injected directly into the circulation of 72 hpf *Tg(kdrl:HRAS:mCherry)* larvae that previously showed haemorrhage, which was mostly localized around the hindbrain ventricle. Confocal micrographs

of 72 hpf *Tg(kdrl:HRAS:mCherry)* expression, FITC-dextran and MERGE of WT and *egfl7*^{981/s981} larvae in (a) lateral and (b) dorsal views. The FITC-dextran did not accumulate to the sites of haemorrhage, suggesting that these sites had clotted and vascular integrity had been restored after the initial blood leakage.

a

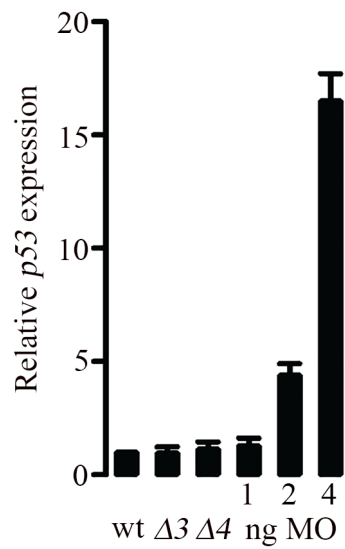


b

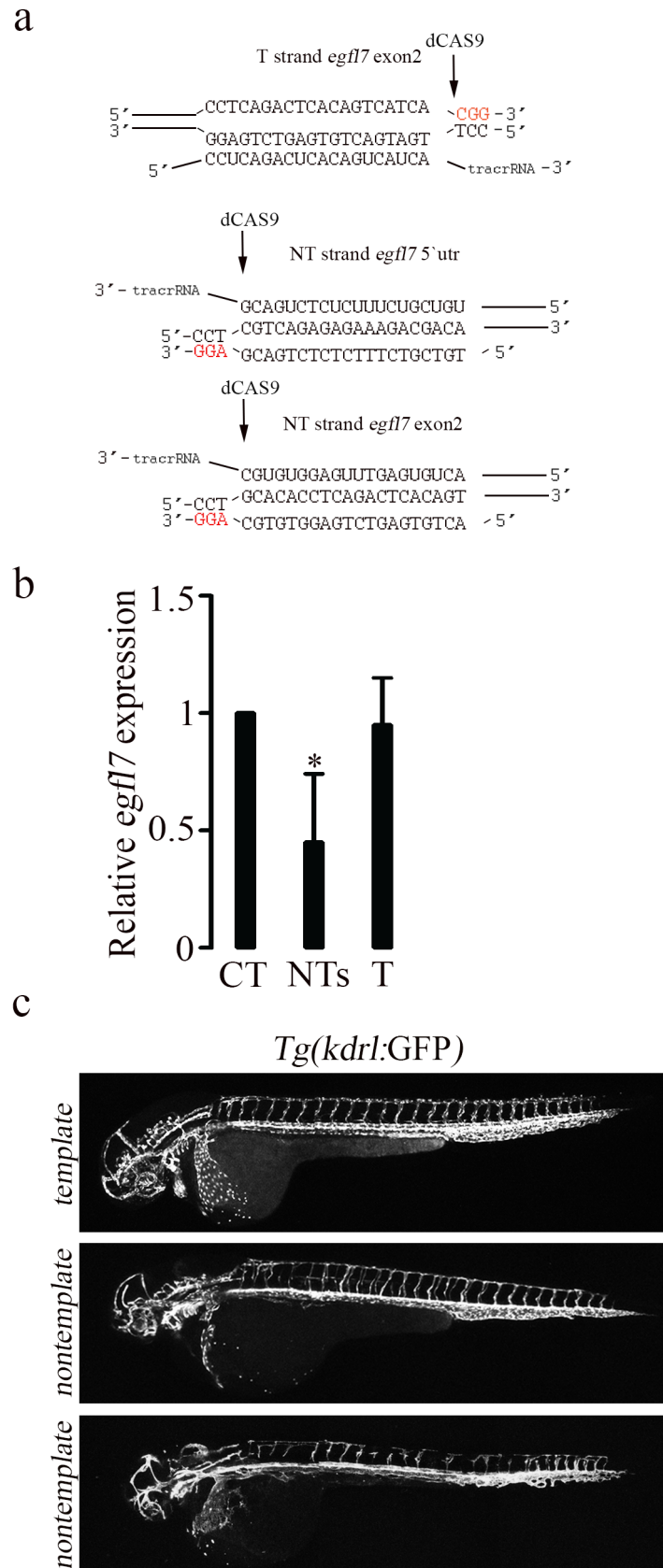


Extended Data Figure 4 | *In vivo* genome editing: Myc-tag introduction in the *egfl7* endogenous locus. **a**, TALENs targeting the *egfl7* stop codon created double-stranded breaks in the chromosomal DNA. Homology-directed repair precisely incorporated the Myc tag exogenous sequence (ssDNA) at the

cut site. **b**, Western blot analysis of Egfl7-Myc-tag expression in 24 hpf control and morphant embryos. Egfl7 Myc-tag signal was reduced by around 80% in morphants (1 ng *egfl7* MO) compared with uninjected. Expression normalized to tubulin ($P \leq 0.05$). Error bars, s.e.m. ($n = 3$).



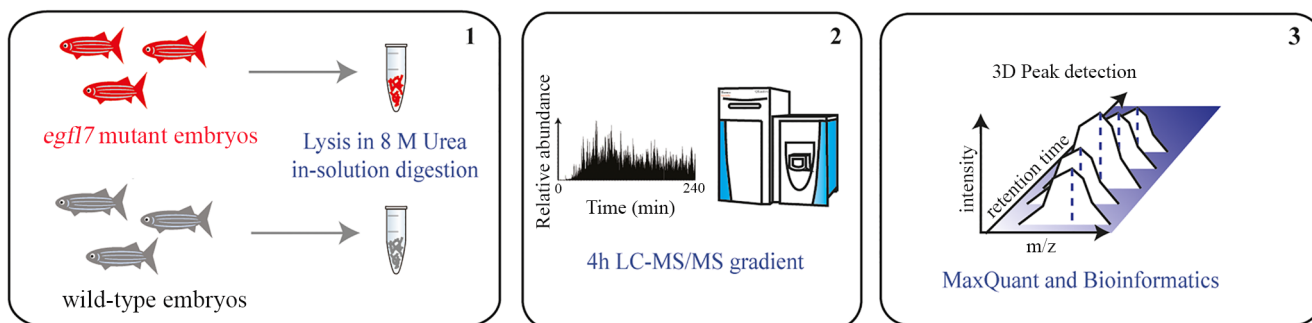
Extended Data Figure 5 | The *egfl7* morpholino does not significantly affect *p53* mRNA expression at 1 ng per embryo but it does so at higher doses. mRNA expression of *p53* in 24 hpf WT, *egfl7* $\Delta 3$ (*egfl7*⁹⁸⁰) and *egfl7* $\Delta 4$ (*egfl7*⁹⁸¹) mutant, and morphant (1, 2 and 4 ng injected) embryos. Expression normalized to *gapdh*. Error bars, s.e.m. of technical triplicates.



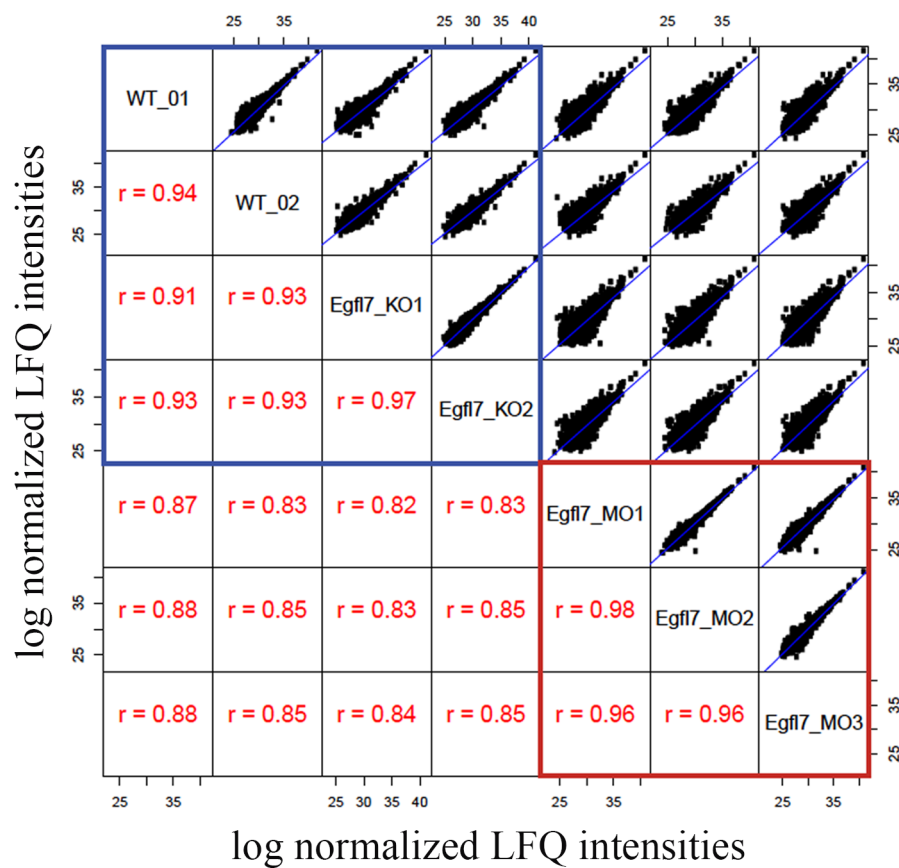
Extended Data Figure 6 | The *egfl7* transcript elongation inhibition causes a phenotype similar to the one seen in morphants. a, gRNAs of *egfl7* targeting the template (T) strand in exon 2 and non-template (NT) strand in the 5' UTR and exon 2. **b,** Expression of *egfl7* in non-template (NT) gRNA and template (T) gRNA-injected embryos relative to uninjected (CT) siblings at 20 hpf. qPCR data, pools of ten embryos each, expression normalized to *gapdh*

($P \leq 0.05$). Error bars, s.e.m. ($n = 3$). **c,** Lateral view confocal micrographs of 48 hpf *Tg(kdrl:GFP)* embryos injected with *egfl7* template and non-template CRISPRi. Template CRISPRi (top) embryos are indistinguishable from non-injected siblings, while non-template CRISPRi embryos exhibit different degrees of vascular defects (middle: mild; bottom: severe).

a



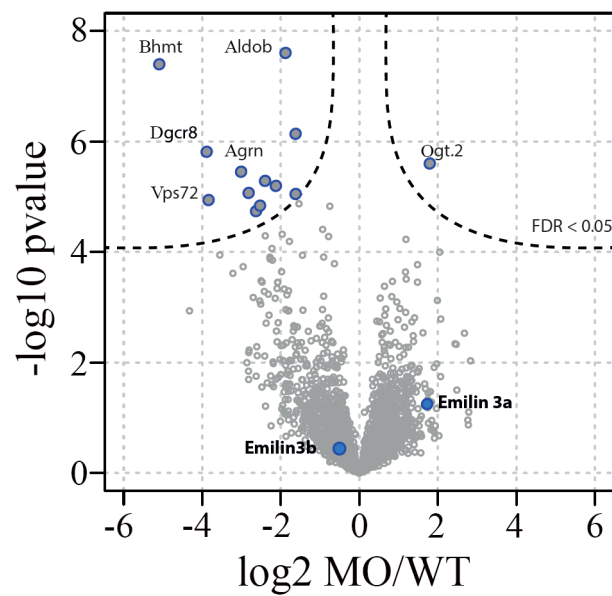
b



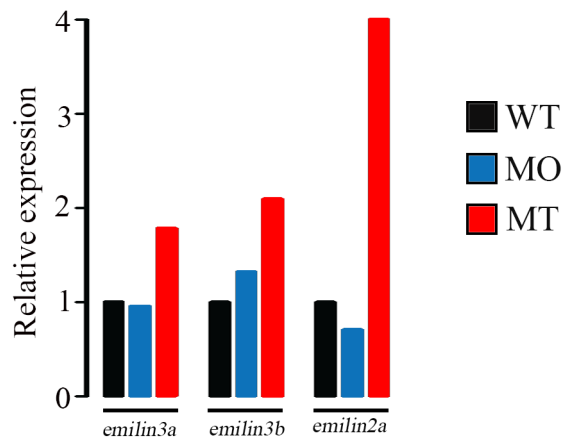
Extended Data Figure 7 | Single-shot proteomics to assess changes between WT and *egfl7*⁹⁸¹ mutant embryos. **a**, Schematic visualization of proteomic workflow. Embryos were lysed in urea buffer, and proteins were digested in-solution using trypsin and measured on a QExactive bench top instrument.

Acquired spectra were analysed against the Uniprot zebrafish database (2014) using MaxQuant. **b**, Scatter plot matrix shows high correlation between biological replicates. Reproducibility was determined by a Pearson correlation coefficient.

a



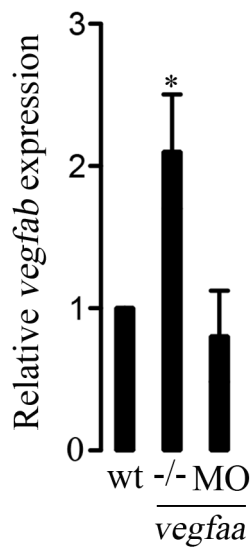
b



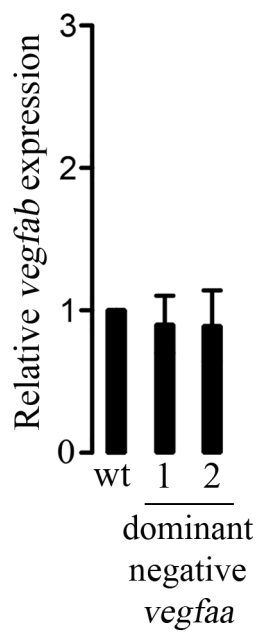
Extended Data Figure 8 | Emilin3a expression is upregulated in mutant but not morphant embryos. a, Volcano plot showing significantly dysregulated proteins between *egfl7* morphant and WT embryos at 24 hpf using label-free quantification. Emilin3a (blue) levels were not significantly different

between morphant and WT embryos. Emilin3b is also highlighted in blue. b, Bar plot showing upregulation of *emilin* family members in 24 hpf *egfl7* mutants compared with WT and morphants, as assessed from RNA-seq data (WT expression set at 1 for each gene).

a



b



Extended Data Figure 9 | Expression of *vegfab* is upregulated in *vegfaa* mutant embryos but not in morphants, or *vegfaa* dominant negative-injected embryos; qPCR data, pools of ten embryos each, expression normalized to *gapdh* ($P \leq 0.05$). Error bars, s.e.m. ($n = 5$). **a, mRNA expression of *vegfab* in 24 hpf *vegfaa* WT, mutant and morphant embryos. **b**, mRNA expression of *vegfab* in 24 hpf *vegfaa* WT and *vegfaa* dominant negative-injected embryos (two different dominant negatives were injected).**

Metabolic rescue in pluripotent cells from patients with mtDNA disease

Hong Ma^{1,2}, Clifford D. L. Folmes³, Jun Wu⁴, Robert Morey⁵, Sergio Mora-Castilla⁵, Alejandro Ocampo⁴, Li Ma⁴, Joanna Poulton⁶, Xinjian Wang⁷, Riffat Ahmed^{1,2}, Eunju Kang^{1,2}, Yeonmi Lee^{1,2}, Tomonari Hayama^{1,2}, Ying Li^{1,2}, Crystal Van Dyken^{1,2}, Nuria Marti Gutierrez^{1,2}, Rebecca Tippner-Hedges^{1,2}, Amy Koski^{1,2}, Nargiz Mitalipov^{1,2}, Paula Amato⁸, Don P. Wolf², Taosheng Huang⁷, Andre Terzic³, Louise C. Laurent⁵, Juan Carlos Izpisua Belmonte⁴ & Shoukhrat Mitalipov^{1,2}

Mitochondria have a major role in energy production via oxidative phosphorylation¹, which is dependent on the expression of critical genes encoded by mitochondrial (mt)DNA. Mutations in mtDNA can cause fatal or severely debilitating disorders with limited treatment options². Clinical manifestations vary based on mutation type and heteroplasmy (that is, the relative levels of mutant and wild-type mtDNA within each cell)^{3,4}. Here we generated genetically corrected pluripotent stem cells (PSCs) from patients with mtDNA disease. Multiple induced pluripotent stem (iPS) cell lines were derived from patients with common heteroplasmic mutations including 3243A>G, causing mitochondrial encephalomyopathy and stroke-like episodes (MELAS)⁵, and 8993T>G and 13513G>A, implicated in Leigh syndrome. Isogenic MELAS and Leigh syndrome iPS cell lines were generated containing exclusively wild-type or mutant mtDNA through spontaneous segregation of heteroplasmic mtDNA in proliferating fibroblasts. Furthermore, somatic cell nuclear transfer (SCNT) enabled replacement of mutant mtDNA from homoplasmic 8993T>G fibroblasts to generate corrected Leigh-NT1 PSCs. Although Leigh-NT1 PSCs contained donor oocyte wild-type mtDNA (human haplotype D4a) that differed from Leigh syndrome patient haplotype (F1a) at a total of 47 nucleotide sites, Leigh-NT1 cells displayed transcriptomic profiles similar to those in embryo-derived PSCs carrying wild-type mtDNA, indicative of normal nuclear-to-mitochondrial interactions. Moreover, genetically rescued patient PSCs displayed normal metabolic function compared to impaired oxygen consumption and ATP production observed in mutant cells. We conclude that both reprogramming approaches offer complementary strategies for derivation of PSCs containing exclusively wild-type mtDNA, through spontaneous segregation of heteroplasmic mtDNA in individual iPS cell lines or mitochondrial replacement by SCNT in homoplasmic mtDNA-based disease.

Maternally inherited mtDNA encodes 13 proteins critical for oxidative phosphorylation, while the remaining protein subunits are encoded by nuclear DNA. Therefore, mitochondrial biogenesis requires coordinated interaction of protein subunits encoded by both genomes². Mutations in mtDNA occur at a higher rate than in nuclear DNA, resulting in life-threatening conditions^{3,6}.

We have described a strategy to prevent transmission of mtDNA mutations to children involving mitochondrial replacement⁷. To explore the feasibility of generating genetically corrected autologous PSCs, herein, we focus on three of the most common pathogenic mtDNA mutations. Skin samples were donated by a MELAS patient carrying a 3243A>G heteroplasmic mutation in tRNA^{Leu} (*MT-TL1*)⁸

and by Leigh syndrome patients carrying heteroplasmic or homoplasmic 8993T>G mutations affecting the ATPase 6 gene (*MT-ATP6*)³, and heteroplasmic 13513G>A mutation in the *MT-ND5* gene⁹. A panel of ten iPS cell lines from each mutation type was generated and quantitative mtDNA mutation analysis was carried out using amplification refractory mutation system-quantitative polymerase chain reaction (ARMS-qPCR), with a detection threshold of 0.5%. In MELAS iPS cell lines, the mutation was undetectable in five lines and varied from 33% to 100% in the remaining five lines, compared to 29% heteroplasmy in parental fibroblasts (Table 1 and Extended Data Fig. 1a). In iPS cell lines from the heteroplasmic 8993T>G mutation, the mutation was undetectable in one line and ranged from 29% to 87% in the remaining lines, compared to 52% heteroplasmy in parental fibroblasts (Table 1 and Extended Data Fig. 1b). Mutation segregation in individual iPS cell lines from 13513G>A fibroblasts also ranged from 0% to 100%, compared to 84% heteroplasmy in fibroblasts (Table 1 and Extended Data Fig. 1c). Previous studies suggested that segregation of heteroplasmic mtDNA is specific to iPS cells and may occur during or after reprogramming^{10,11}. To explore mechanisms, parental fibroblasts carrying 3243A>G and 13513G>A mutations were subcloned and mutation loads in individual clones were analysed. Among ten randomly selected MELAS samples, five were homoplasmic containing either wild type (A) or mutant (G) at the 3243 position. The remaining five contained varying heteroplasmy levels similar to iPS cells (Table 1 and Extended Data Fig. 1d). Variable heteroplasmy levels were also observed in 13513G>A fibroblasts including homoplasmic mutant and wild-type clones (Table 1). Thus, segregation of heteroplasmic mtDNA mutations occurs in skin fibroblasts and may reflect a common phenomenon¹².

Isogenic MELAS iPS cell lines carrying wild-type or mutant mtDNA maintained typical PSC morphology and formed teratomas containing cells and tissues from all three germ layers (Extended Data Fig. 2a, b). We next carried out whole mtDNA sequencing using the Illumina MiSeq platform and confirmed the 3243A>G mutation in parental MELAS-fib (46.8%), MELAS-iPS1 and MELAS-iPS3 (100%) cells while MELAS-iPS2 was homoplasmic for the wild-type allele (Supplementary Table 1). MELAS-fib also carried four additional heteroplasmic mutations with one variant carried to MELAS-iPS1 and MELAS-iPS2 (Supplementary Table 1).

The 3243A>G mutation perturbs tRNA^{Leu} function and impairs mitochondrial protein synthesis as well as respiratory complex activity, with the homoplasmic mutation leading to prenatal lethality in humans¹³. Oxygen consumption rate (OCR) was employed as an indicator of mitochondrial respiration and energy production. Mutant MELAS-iPS1 and MELAS-iPS3 exhibited significantly lower OCR

¹Center for Embryonic Cell and Gene Therapy, Oregon Health & Science University, 3303 S.W. Bond Avenue, Portland, Oregon 97239, USA. ²Division of Reproductive & Developmental Sciences, Oregon National Primate Research Center, Oregon Health & Science University, 505 N.W. 185th Avenue, Beaverton, Oregon 97006, USA. ³Center for Regenerative Medicine and Department of Medicine, Division of Cardiovascular Diseases, Mayo Clinic, Rochester, Minnesota 55905, USA. ⁴Gene Expression Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. ⁵Department of Reproductive Medicine, University of California, San Diego, Sanford Consortium for Regenerative Medicine, 2880 Torrey Pines Scenic Drive, La Jolla, California 92037, USA. ⁶Department of Obstetrics and Gynaecology, John Radcliffe Hospital, University of Oxford, Headington, Oxford OX3 9DU, UK. ⁷Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. ⁸Division of Reproductive Endocrinology, Department of Obstetrics and Gynecology, Oregon Health and Science University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239, USA.

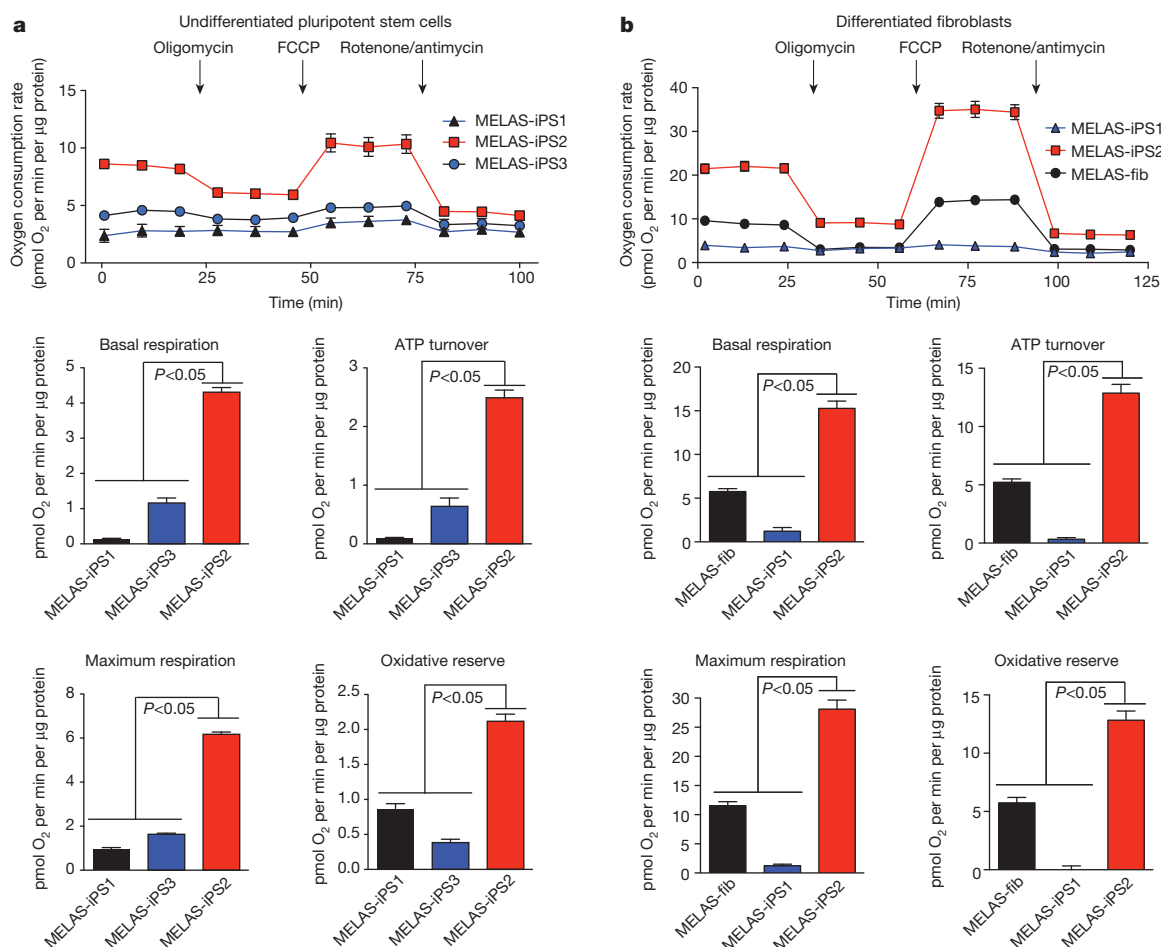
Table 1 | Distribution of mtDNA variants in fibroblast and iPS clones derived from patients with heteroplasmic mutations

iPS clones						Fibroblast clones			
3243 A>G	Mutant G (%)	8993 T>G	Mutant G (%)	13513 G>A	Mutant A (%)	3243 A>G	Mutant G (%)	13513 G>A	Mutant A (%)
Parental fibroblasts	29	Parental fibroblasts	52	Parental fibroblasts	84	Parental fibroblasts	29	Parental fibroblasts	84
iPS1	100	iPS1	62	iPS1	100	fib1	100	fib1	0
iPS2	0	iPS2	72	iPS2	2	fib2	100	fib2	68
iPS3	100	iPS3	32	iPS3	4	fib3	0	fib3	24
iPS4	0	iPS4	52	iPS4	0	fib4	93	fib4	64
iPS5	0	iPS5	29	iPS5	80	fib5	8	fib5	58
iPS6	33	iPS6	66	iPS6	11	fib6	21	fib6	48
iPS7	0	iPS7	87	iPS7	19	fib7	3	fib7	69
iPS8	78	iPS8	72	iPS8	32	fib8	97	fib8	70
iPS9	88	iPS9	46	iPS9	100	fib9	100	fib9	63
iPS10	0	iPS10	0	iPS10	72	fib10	0	fib10	100

($P < 0.05$) when compared to the wild-type MELAS-iPS2 (Fig. 1a). Fibroblasts differentiated from MELAS-iPS1 and parental MELAS-fib also displayed low levels of mitochondrial function. In contrast, these respiratory defects were absent in MELAS-iPS2-derived fibroblasts. In general, mitochondrial respiration correlated with the heteroplasmy levels in cells (Fig. 1b and Table 1). The greater reliance on oxidative metabolism in wild-type MELAS-iPS2 was confirmed by the elevated OCR to ECAR (extracellular acidification rate) ratio, which provides a measure for the relative contribution of oxidative metabolism versus glycolysis (Extended Data Fig. 3a). Mutant MELAS iPS cells

and their derivatives displayed significantly decreased OCR/ECAR ratios, indicating a greater reliance on glycolysis (Extended Data Fig. 3a, b). We next differentiated MELAS iPS cells into neuronal progenitor cells (NPCs, Extended Data Fig. 3c, d)^{14,15}. Diminished metabolic profiles in mutant NPCs recapitulated those observed in undifferentiated iPS cells (Extended Data Fig. 3e). Cardiomyocyte differentiation¹⁵ of mutant MELAS iPS cells was severely compromised due to massive cell death.

As expected, all iPS cell lines from homoplasmic 8993T>G fibroblasts carried mutant mtDNA (Extended Data Fig. 1e and Extended

**Figure 1 | Mitochondrial respiratory function in MELAS samples.**

a, Oxygen consumption rate (OCR) in undifferentiated MELAS-iPS1, MELAS-iPS2 and MELAS-iPS3 cells ($n = 9$ per cell line, biological replicates) in response to $0.5 \mu\text{g ml}^{-1}$ oligomycin, $1 \mu\text{M}$ fluorocarbonyl cyanide phenylhydrazine (FCCP), $0.5 \mu\text{M}$ rotenone and $1 \mu\text{M}$ antimycin. Wild-type MELAS-iPS2 displayed higher levels of oxygen consumption when compared to mutant

MELAS-iPS1 and MELAS-iPS3. **b**, OCR in MELAS-iPS1 and MELAS-iPS2 derived fibroblasts and parental MELAS-fib ($n = 10$ per cell line, biological replicates). Error bars are mean \pm s.e.m. and OCR data are representative of at least 2–3 independent experiments. Significance established with one-way analysis of variance (ANOVA) with Tukey's multiple comparison test.

Table 2 | Summary of 47 SNPs found in the mtDNA of Leigh-fib and Leigh-NT1 lines

Nucleotide position	Leigh-NT1	Leigh-fib	Locus	Effects	Nucleotide position	Leigh-NT1	Leigh-fib	Locus	Effects
152	C	T	Control region		10400	T	C	<i>MT-ND3</i>	Syn
248	A	Deletion	Control region	–	10410	C	T	<i>MT-TR</i>	–
489	C	T	Control region	–	10609	T	C	<i>MT-ND4L</i>	Non-syn
3010	A	G	<i>MT-RNR2</i>	–	10873	C	T	<i>MT-ND4</i>	Syn
3206	T	C	<i>MT-RNR2</i>	–	12406	G	A	<i>MT-ND5</i>	Non-syn
3970	C	T	<i>MT-ND1</i>	Syn	12418	Deletion	A	<i>MT-ND5</i>	Frame shift
4086	C	T	<i>MT-ND1</i>	Syn	12705	T	C	<i>MT-ND5</i>	Syn
4216	T	C	<i>MT-ND1</i>	Non-syn	12882	C	T	<i>MT-ND5</i>	Syn
4883	T	C	<i>MT-ND2</i>	Syn	13759	G	A	<i>MT-ND5</i>	Non-syn
5178	A	C	<i>MT-ND2</i>	Non-syn	13928	G	C	<i>MT-ND5</i>	Non-syn
6392	T	C	<i>MT-CO1</i>	Syn	14668	T	C	<i>MT-ND6</i>	Syn
6527	A	G	<i>MT-CO1</i>	Syn	14783	C	T	<i>MT-CYB</i>	Syn
6962	G	A	<i>MT-CO1</i>	Syn	14979	C	T	<i>MT-CYB</i>	Non-syn
7775	A	G	<i>MT-CO2</i>	Non-syn	15043	A	G	<i>MT-CYB</i>	Syn
8414	T	C	<i>MT-ATP8</i>	Non-syn	15301	A	G	<i>MT-CYB</i>	Syn
8473	C	T	<i>MT-ATP8</i>	Syn	15676	T	C	<i>MT-CYB</i>	Syn
8507	A	G	<i>MT-ATP8</i>	Non-syn	16148	C	T	Control region	–
8701	G	A	<i>MT-ATP6</i>	Non-syn	16162	A	G	Control region	–
8993	T	G	<i>MT-ATP6</i>	Non-syn	16172	T	C	Control region	–
9053	G	A	<i>MT-ATP6</i>	Non-syn	16223	T	C	Control region	–
9540	C	T	<i>MT-CO3</i>	Syn	16244	G	A	Control region	–
9548	G	A	<i>MT-CO3</i>	Syn	16304	T	C	Control region	–
10310	G	A	<i>MT-ND3</i>	Syn	16362	C	T	Control region	–
10398	G	A	<i>MT-ND3</i>	Non-syn					

Syn, synonymous; non-syn, non-synonymous.

Data Table 1). Therefore, we pursued mitochondrial replacement by SCNT with wild-type oocyte mitochondria. Following our reported protocol, two stable nuclear transfer–embryonic stem cell lines were established (Leigh-NT1 and Leigh-NT2)¹⁶. Genotyping confirmed that both lines contained predominantly oocyte wild-type mtDNA (Extended Data Fig. 1f) with limited low mutated mtDNA carryover (<1%) at passage 5 that became undetectable upon extended propagation (Extended Data Table 2).

Cytogenetic G-banding revealed that Leigh-iPS1 and Leigh-NT1 retained normal diploid karyotypes with no detectable numerical or

structural chromosomal abnormalities (Extended Data Fig. 2c). However, Leigh-NT2 showed a XXXY tetraploid karyotype (Extended Data Fig. 2c). Fingerprinting by short tandem repeat analysis (STR) also revealed that Leigh-NT2 contained both oocyte and Leigh-fib alleles (Extended Data Table 3), consistent with failed enucleation. STR profiles for Leigh-NT1 and Leigh-iPS1 were identical to Leigh-fib (Extended Data Table 3). Both Leigh-iPS1 and Leigh-NT1 lines maintained typical PSC morphology, expressed pluripotency markers⁷ and formed teratomas containing cells and tissues from all three germ layers (Extended Data Fig. 2d, e).

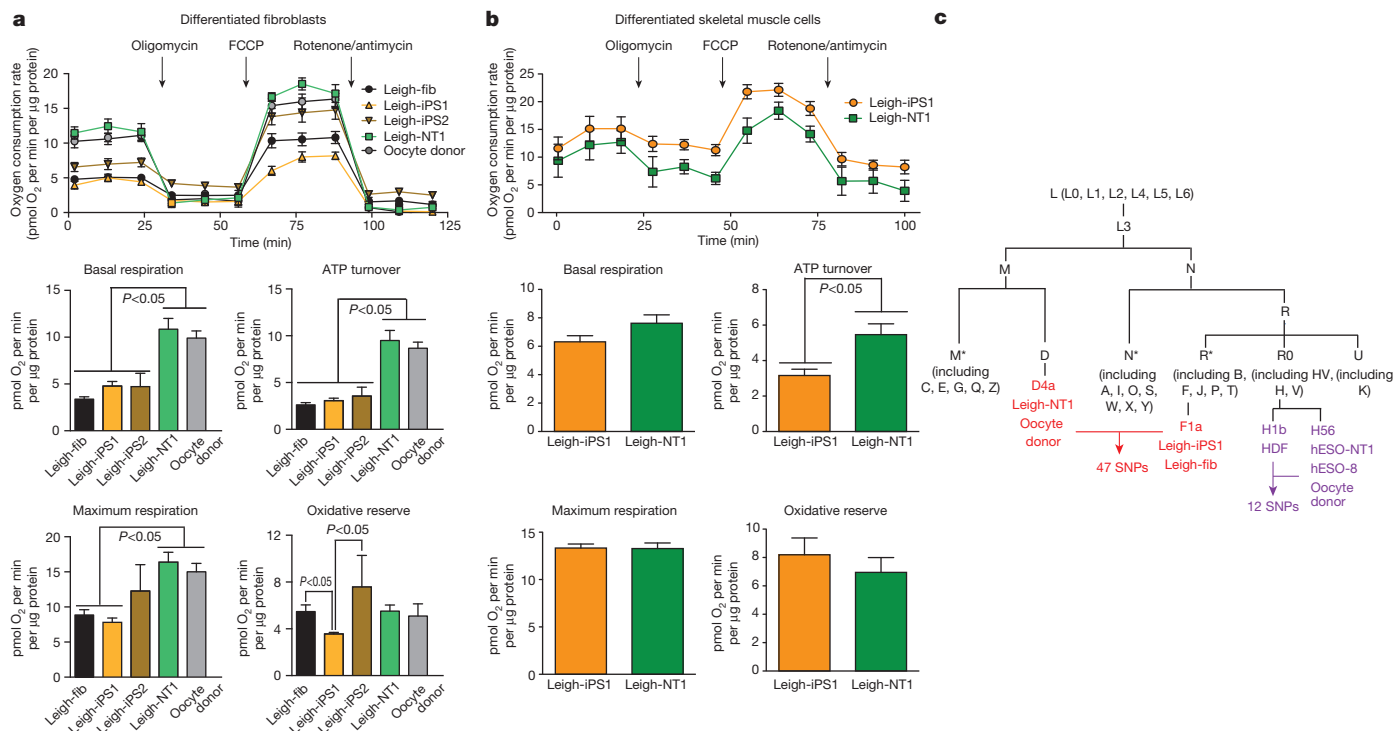


Figure 2 | Restoration of mitochondrial respiratory function in Leigh-NT1. **a**, OCR in Leigh-NT1, Leigh-iPS1 and Leigh-iPS2 derived fibroblasts, parental and oocyte donor fibroblasts ($n = 9, 8, 10, 9$ and 8 per cell line, respectively, biological replicates). **b**, OCR in Leigh-NT1 and Leigh-iPS1 derived skeletal muscle cells ($n = 6$ biological replicates per cell line). **c**, mtDNA haplotype distances for oocyte and somatic cell donors based on mitochondrial

phylogenetic tree from PhyloTree (<http://phylotree.org/tree/main.htm>)²⁷. Asterisks indicate subgroups for mtDNA haplotypes. Error bars are mean \pm s.e.m. and OCR data are representative of at least 2–3 independent experiments. Significance established using Kruskal–Wallis with Dunn’s multiple comparison test or one-way ANOVA with Tukey’s multiple comparison test (maximum respiration in **a**) and Student’s *t*-test in **b**.

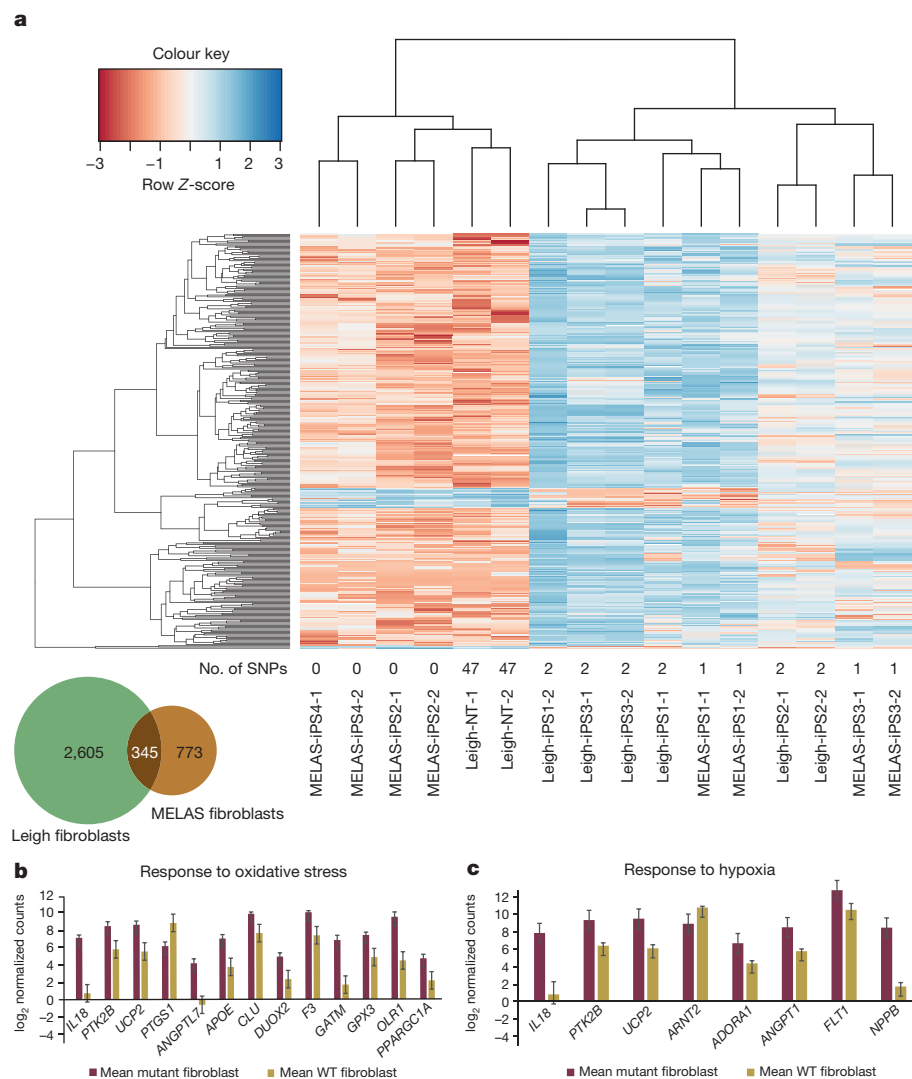


Figure 3 | Global gene expression analysis by RNA-seq. **a**, Heat map displaying 345 genes that are differentially expressed (adjusted P value <0.05) between wild-type PSCs (biological duplicates of 3 independent cell lines) versus mutant PSCs (biological duplicates of 5 independent cell lines). **b**, Functional enrichment analysis of genes displayed in the heat map that are known to be correlated with a response to oxidative stress. **c**, Functional enrichment analysis of genes displayed in the heat map that are known to be correlated with a response to hypoxia. Bar graphs are mean \pm s.e.m., using all samples described in **a**.

Whole mtDNA sequencing confirmed the presence of homoplasmic 8993T>G mutation in Leigh-fib and Leigh-iPS1 and also identified a second homoplasmic 4216T>C mutation in the *MT-ND1* gene (Extended Data Fig. 1f and Supplementary Table 2). This non-synonymous mutation has been previously associated with Leber's hereditary optic neuropathy¹⁵. Leigh-NT1 mtDNA sequence differed from Leigh-fib at 47 nucleotide positions. In addition to the pathogenic 8993T>G and 4216T>C mutations, differences included 10 single nucleotide polymorphisms (SNPs) in the D-loop region, 2 in the 16S rRNA gene, 1 in the tRNA-R gene and 34 in protein genes (Table 2). We also detected two heteroplasmic variants in Leigh-NT1, four in Leigh-fib and three in Leigh-iPS1 (Supplementary Table 2). Clinical symptoms associated with these variants have not been reported.

We measured metabolic function in fibroblasts differentiated from Leigh-NT1, Leigh-iPS1 and Leigh-iPS2 and compared them to parental Leigh-fib and healthy skin fibroblasts from the oocyte donor. As expected, the homoplasmic 8993T>G mutation resulted in low mitochondrial oxidative capacity. In contrast, these respiration defects were absent in fibroblasts differentiated from Leigh-NT1 (Fig. 2a). Leigh-NT1 also displayed a metabolic profile and OCR/ECAR ratios similar to oocyte donor fibroblasts (Fig. 2a and Extended Data Fig. 4a). We observed varying levels of oxidative reserve for Leigh-iPS1 and Leigh-iPS2 compared to parental Leigh-fib (Fig. 2a), reflecting inherent variability within differentiated fibroblast populations. Both Leigh-NT1 and Leigh-iPS1 effectively generated skeletal muscle cells¹⁷ (Extended Data Fig. 4b), with Leigh-iPS1 skeletal muscle cells

displaying significantly lower ATP turnover ($P < 0.05$) (Fig. 2b). Extensive cell death was observed in Leigh-iPS1 during directed cardiomyocyte differentiation (Extended Data Fig. 4c). These results demonstrate complete functional rescue of mitochondrial activity in Leigh-NT1 through restoration of the wild-type mtDNA.

Evolution of mtDNA has resulted in a series of neutral polymorphic variants within the human population often associated with regional migration and adaptation to climate¹⁸. The largest difference between distant human mtDNA haplotypes has been estimated at 95 SNPs¹⁹. In the present study, phylogenetic analysis assigned oocyte and Leigh-NT1 mtDNA to the D4a haplotype while the Leigh-iPS1 and Leigh-fib mtDNA haplotype was F1a (Supplementary Table 2). D4a is a descendant from the M while F1a comes from the N macro-haplo-group per the human mtDNA mutation tree (Fig. 2c). Safety evaluations of mitochondrial replacement therapy suggest possible harmful secondary outcomes reflecting nuclear-mitochondrial incompatibility^{20,21}. Despite 'unmatched' donor mtDNA, Leigh-NT1 demonstrated lineage-specific differentiation and restoration of metabolic activity, implying normal nuclear-mitochondrial interaction. We further investigated a hESO-NT1 derived by SCNT from healthy fetal fibroblasts (human dermal fibroblast (HDF)) and IVF-derived hESO-8 carrying identical mtDNA¹⁶. hESO-NT1 mtDNA differed from HDF at 12 nucleotide positions (Fig. 2c and Supplementary Table 3). Metabolic profiles in NPCs and cardiomyocytes differentiated from hESO-NT1 and hESO-8 displayed similar metabolic profiles (Extended Data Fig. 5a–e). Next, we asked whether the 3243A>G

and 8993T>G mutations induced detectable changes in global gene expression and compared transcriptomes by RNA-seq for undifferentiated and differentiated PSCs. Undifferentiated PSCs containing wild-type or mutant mtDNA showed 154 differentially expressed transcripts (adjusted P value <0.05). This small number of differences is consistent with the predominantly glycolytic metabolism of pluripotent stem cells, which protects them from the deleterious effects of mtDNA mutations²². Global gene expression analysis of fibroblasts differentiated from isogenic MELAS lines identified 1,118 differentially expressed genes in mutant and wild-type cells (Extended Data Fig. 6a), whereas 2,950 genes were differentially expressed in fibroblasts differentiated from mutant Leigh-iPS1, Leigh-iPS2 and Leigh-iPS3 compared to the wild-type Leigh-NT1 cells (Extended Data Fig. 6b). Hierarchical clustering using a multiple bootstrap resampling algorithm showed that the Leigh-NT1 fibroblasts were similar to hESO-NT1, hESO-NT2, hESO-7 and hESO-8 fibroblasts (Extended Data Fig. 6c). These findings further support the notion that oocyte mtDNA in Leigh-NT1 interacts normally with nuclear DNA as long as the mtDNA sequence differences are neutral.

Next, we asked whether any of the differentially expressed genes were common to both 3243A>G and 8993T>G mutations, and found 345 genes that were shared, 96% of which were overexpressed in the mutant cells (Fig. 3a). Functional enrichment analysis identified genes associated with a response to hypoxia and oxidative stress^{23,24} (Fig. 3b, c; P value <0.001). However, we did not observe an enrichment of genes associated with metabolism, stress response, epigenetic regulation, and cell signalling, which was reported in a recent MELAS study¹³ (Extended Data Fig. 6d).

Finally, we addressed whether the 3243A>G and 8993T>G mutations specifically impact gene expression of mtDNA-encoded transcripts²⁵. We found that transcripts expressed from mtDNA accounted for approximately 20% of the total cellular transcriptome, with similar expression levels across different mutations (Extended Data Fig. 7; adjusted P value >0.05).

We demonstrate complementary strategies for generating genetically and functionally corrected PSCs for patients with mtDNA disease. For the most common mtDNA syndromes caused by heteroplasmic mutations, generation of multiple iPS cell lines allows recovery of clones with exclusively wild-type mtDNA due to spontaneous segregation of heteroplasmic mtDNA. SCNT enables correction of homoplasmic mutations through replacement with donor mtDNA, and generation of PSCs with transcriptional and epigenetic profiles similar to embryo-derived embryonic stem cells²⁶. Recovery of metabolic function despite haplotype differences between patient and donor mtDNA suggests that normal nuclear-to-mitochondrial interactions are highly conserved within species. Generation of genetically corrected PSCs from patients with mtDNA disease enables the transition from palliative care to therapeutic interventions based on regenerative medicine.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 September 2014; accepted 12 May 2015.

Published online 15 July 2015.

- McBride, H. M., Neuspiel, M. & Wasiak, S. Mitochondria: more than just a powerhouse. *Curr. Biol.* **16**, 551–560 (2006).
- Wallace, D. C. A mitochondrial bioenergetic etiology of disease. *J. Clin. Invest.* **123**, 1405–1412 (2013).
- Taylor, R. W. & Turnbull, D. M. Mitochondrial DNA mutations in human disease. *Nature Rev. Genet.* **6**, 389–402 (2005).
- Grossman, L. I. & Shoubridge, E. A. Mitochondrial genetics and human disease. *BioEssays* **18**, 983–991 (1996).
- Goto, Y., Nonaka, I. & Horai, S. A mutation in the tRNA^{Leu(UUR)} gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature* **348**, 651–653 (1990).
- Tuppen, H. A., Blakely, E. L., Turnbull, D. M. & Taylor, R. W. Mitochondrial DNA mutations and human disease. *Biochim. Biophys. Acta* **1797**, 113–128 (2010).

- Tachibana, M. *et al.* Towards germline gene therapy of inherited mitochondrial diseases. *Nature* **493**, 627–631 (2013).
- Morten, K. J., Poulton, J. & Sykes, B. Multiple independent occurrence of the 3243 mutation in mitochondrial tRNA^{Leu(UUR)} in patients with the MELAS phenotype. *Hum. Mol. Genet.* **4**, 1689–1691 (1995).
- Choi, M. *et al.* The mitochondrial DNA G13513A MELAS mutation in the NADH dehydrogenase 5 gene is a frequent cause of Leigh-like syndrome with isolated complex I deficiency. *J. Med. Genet.* **40**, 188–191 (2003).
- Fujikura, J. *et al.* Induced pluripotent stem cells generated from diabetic patients with mitochondrial DNA A3243G mutation. *Diabetologia* **55**, 1689–1698 (2012).
- Hamalainen, R. H. *et al.* Tissue- and cell-type-specific manifestations of heteroplasmic mtDNA 3243A>G mutation in human induced pluripotent stem cell-derived disease model. *Proc. Natl Acad. Sci. USA* **110**, E3622–E3630 (2013).
- Folmes, C. D. *et al.* Disease-causing mitochondrial heteroplasmy segregated within induced pluripotent stem cell clones derived from a patient with MELAS. *Stem Cells* **31**, 1298–1308 (2013).
- Picard, M. *et al.* Progressive increase in mtDNA 3243A>G heteroplasmy causes abrupt transcriptional reprogramming. *Proc. Natl Acad. Sci. USA* **111**, E4033–E4042 (2014).
- Liu, G. H. *et al.* Progressive degeneration of human neural stem cells caused by pathogenic LRRK2. *Nature* **491**, 603–607 (2012).
- Lian, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/ β -catenin signaling under fully defined conditions. *Nature Protocols* **8**, 162–175 (2013).
- Tachibana, M. *et al.* Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell* **153**, 1228–1238 (2013).
- Borichin, B., Chen, J. & Barberi, T. Derivation and FACS-mediated purification of PAX3⁺/PAX7⁺ skeletal muscle precursors from human pluripotent stem cells. *Stem Cell Rev.* **1**, 620–631 (2013).
- Wallace, D. C., Brown, M. D. & Lott, M. T. Mitochondrial DNA variation in human evolution and disease. *Gene* **238**, 211–230 (1999).
- Brandon, M. C. *et al.* MITOMAP: a human mitochondrial genome database–2004 update. *Nucleic Acids Res.* **33**, D611–D613 (2005).
- Burgstaller, J. P. *et al.* MtDNA segregation in heteroplasmic tissues is common *in vivo* and modulated by haplotype differences and developmental stage. *Cell Rep.* **7**, 2031–2041 (2014).
- Chinnery, P. F. *et al.* The challenges of mitochondrial replacement. *PLoS Genet.* **10**, e1004315 (2014).
- Folmes, C. D., Dzeja, P. P., Nelson, T. J. & Terzic, A. Metabolic plasticity in stem cell homeostasis and differentiation. *Cell Stem Cell* **11**, 596–606 (2012).
- Henchcliffe, C. & Beal, M. F. Mitochondrial biology and oxidative stress in Parkinson disease pathogenesis. *Nat. Clin. Pract. Neurol.* **4**, 600–609 (2008).
- Boland, M. L., Chourasia, A. H. & Macleod, K. F. Mitochondrial dysfunction in cancer. *Front. Oncol.* **3**, 292 (2013).
- Mercer, T. R. *et al.* The human mitochondrial transcriptome. *Cell* **146**, 645–658 (2011).
- Ma, H. *et al.* Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature* **511**, 177–183 (2014).
- van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors acknowledge the OHSU Embryonic Stem Cell Research Oversight Committee and the Institutional Review Board for providing oversight and guidance. We thank skin and oocyte donors and the Women's Health Research Unit staff at the Center for Women's Health, University Fertility Consultants and the Reproductive Endocrinology & Infertility Division in the Department of Obstetrics & Gynecology of Oregon Health & Science University for their support and procurement of gametes. We are grateful to M. Tachibana and A. Polat for help with derivation of PSCs and to M. Sparman for technical support. We are indebted to S. Gokhale for teratoma analysis and M. C. T. Penedo for microsatellite genotyping. We thank the staff at the Institute for Genomic Medicine Genomics Facility at UCSD for sequencing the RNA-seq libraries. Studies were supported by the Leducq Foundation, Mayo Clinic Center for Regenerative Medicine and OHSU and UCSD institutional funds. Work in the laboratory of J.C.I.B. was supported by the G. Harold and Leila Y. Mathers Charitable Foundation and the Leona M. and Harry B. Helmsley Charitable Trust (2012-PG-MED002).

Author Contributions H.M. and S.M. conceived the study and designed the experiments. S.M., P.A., H.M., R.A., E.K., Y.L., N.M.G. and R.T.-H. derived and cultured PSCs. J.P. derived MELAS fibroblasts. H.M., T.H., Y.L., C.V.D., A.K. and N.M. performed the DNA/RNA extractions and mtDNA ARMS-qPCR analyses. C.D.L.F. and A.T. performed Seahorse studies on differentiated fibroblasts and data analysis. X.W. and T.H. performed MiSeq studies. R.M., S.M.-C. and L.C.L. performed RNA-seq and bioinformatic analysis of the data. J.W., A.O., L.M. and J.C.I.B. performed NPC, skeletal muscle cell, and cardiomyocyte differentiations, Seahorse studies and data analysis. H.M., C.D.L.F., J.W., R.M., D.P.W., L.C.L., A.T., J.C.I.B. and S.M. analysed the data and wrote the paper.

Author Information Processed data sets can be downloaded from NCBI GEO under accession GSE61390 for RNA-seq. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M. (mitalipo@ohsu.edu).

METHODS

The study protocols and informed consent for human subjects were approved by the OHSU Embryonic Stem Cell Research Oversight Committee and the Institutional Review Board. No statistical methods were used to predetermine sample size.

SCNT and iPS cell derivation and culture. Fibroblasts were acquired from Coriell Cell Repositories or donated by patients directly for our study. Fibroblasts were cultured in DMEM F12 medium supplemented with 10% fetal bovine serum (HyClone) and 50 μ M uridine. SCNT procedures were performed as described previously¹⁶. Sendai virus-based reprogramming was carried out according to the manufacturer's protocol (CytoTune-iPS Reprogramming Kit, Life Technologies). Colonies with typical iPS cell morphology were isolated and manually propagated as described previously²⁶ in Knockout DMEM medium (Invitrogen) supplemented with 20% knockout serum replacement (Invitrogen), 0.1 mM nonessential amino acids (Invitrogen), 1 mM L-glutamine (Invitrogen), 0.1 mM β -mercaptoethanol (Sigma), 1 \times penicillin–streptomycin (Invitrogen) and 4 ng ml⁻¹ basic fibroblast growth factor (Sigma). All cell cultures were free of mycoplasma contamination. Origin of all cell lines has been authenticated by STR and mtDNA genotyping.

Fibroblast differentiation. Differentiation of PSCs to fibroblasts was induced by culture in fibroblast medium (DMEM F12 with 10% FBS) for 2–3 weeks in absence of mouse embryonic fibroblast (mEF) feeder layers. Resulting differentiated cells were FACS sorted for TRA-1-60⁻ (BD Biosciences), SSEA4⁻ (Santa Cruz), CD56⁻ (BD Biosciences) and CD13⁺ (BD Biosciences) cells²⁸. The CD13⁺ cells were further expanded in the fibroblast medium.

NPC differentiation and culture. For NPC differentiation, a published protocol¹⁴ was followed with minor modifications. PSCs were collected using collagenase IV (Life Technologies), washed twice with 1 \times DPBS without calcium and magnesium (Corning Cellgro), and cultured in Neural Induction Medium 1 (NIM-1: 50% Advanced DMEM/F12 (Invitrogen), 50% Neurobasal (Invitrogen), 1 \times B27 (Invitrogen), 1 \times N2 (Invitrogen), 2 mM GlutaMAX (Invitrogen) supplemented with 10 ng ml⁻¹ hLIF (Peprotech), 4 μ M CHIR99021 (Selleckchem), 3 μ M SB431542 (Selleckchem), 2 μ M dorsomorphin (Sigma), and 0.1 μ M Compound E (EMD Chemicals Inc.)). Cells were cultured in NIM-1 medium for 2 days with daily medium change and then switched to Neural Induction Medium 2 (NIM-2: 50% Advanced DMEM/F12, 50% Neurobasal, 1 \times N2, 1 \times B27, 2 mM GlutaMAX and 10 ng ml⁻¹ hLIF, 4 μ M CHIR99021, 3 μ M SB431542 and 0.1 μ M Compound E). After 5 days culture in NIM-2 (daily medium change), cells were treated with 10 μ M Y27632 (Selleckchem) for 1 h and 'dome'-shaped colonies were manually picked and treated with Accumax (Innovative Cell Technologies) for 10 min at 37 °C. Cells were then gently pipetted to obtain single cell suspension and replated onto Matrigel-coated 6-well plates at a density 3.5×10^4 per cm² in Neural Progenitor cell Maintenance Medium (NPM: 50% Advanced DMEM/F12, 50% Neurobasal, 1 \times B27, 1 \times N2, 2 mM GlutaMAX, 10 ng ml⁻¹ hLIF, 3 μ M CHIR99021 and 2 μ M SB431542) supplemented with 10 μ M Y27632. NPCs were maintained on Matrigel-coated dishes in NPM with daily medium change and passaged upon reaching 70% to 80% confluence using Accumax.

Skeletal muscle differentiation. Skeletal muscle differentiation was based on a previous report with minor modifications¹⁷. Briefly, PSCs plated on Matrigel-coated plates were grown to 40% confluence in mTeSR1 medium and then switched to Skeletal Muscle Induction Medium (SMIM: DMEM/F12, ITS, 3 μ M CHIR99021). After 4 days culture in SMIM with daily medium change, cells were cultured in Skeletal Muscle Expansion Medium (SME: DMEM/F12, ITS and 20 ng ml⁻¹ FGF2) for an additional 14 days with daily medium change. Cells were then cultured in Skeletal Muscle Differentiation Medium (SMDM: DMEM/F12 and ITS only) for an additional 18 days.

Cardiomyocyte differentiation. Cardiomyocyte differentiation was performed with adaptation based on the inhibition of GSK3 and Wnt pathways¹⁵. Briefly, PSCs were collected after Accutase (Life Technologies) treatment and cultured on Matrigel-coated plates in RPMI supplemented with B27 without insulin (Invitrogen) to 80–90% confluency. Cells were then incubated with 12 μ M CHIR99021 (Selleckchem) for 16 h. At day 3, cells were incubated with 5 μ M IWP2 (Tocris) for 48 h. At day 7, medium was replaced to RPMI supplemented with complete B27. Medium was replaced every 3 days. Contracting cardiomyocytes were observed on day 12 of differentiation.

Immunocytochemistry. Cultured cells were fixed with 4% paraformaldehyde for 15 min at room temperature and then permeabilized with 0.2% Triton X-100 in PBS for 10 min. Cells were washed 3 \times with PBST (PBS 1 \times , 0.02% Tween-20) and blocked with 10% goat or donkey serum (Sigma) for 1 h at room temperature. Cells were then incubated with primary antibodies diluted in PBST overnight at 4 °C, washed 3 \times with PBST and incubated with secondary antibodies (1:500, Molecular Probes) for 1 h at room temperature. Cells were washed 3 \times and mounted in

Prolong Gold Antifade Mountant (Life Technologies). Image acquisition was performed on a Zeiss LSM 780 confocal microscope. Primary antibodies were: PAX6 (1:100, Convance), NESTIN (1:200, Millipore), MF20 (1:100, DSHB), OCT4 (1:100, Santa Cruz) and NANOG (1:40, R&D Systems).

Teratoma assay. Approximately 3–5 million undifferentiated PSCs were injected into the hindleg muscle of 8-week-old, SCID, beige male mice (Charles River) using an 18-gauge needle. Six to seven weeks after injection, mice were euthanized and tumours were dissected, sectioned and histologically characterized for the presence of representative tissues as described previously⁷. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

mtDNA heteroplasmy analysis by ARMs-qPCR. The amplification refractory mutation system quantitative PCR assay (ARMs-qPCR) was used to measure mtDNA carryover in Leigh-NT1 and NT2 as previously described¹⁶. Primers and TaqMan MGB probes were designed to detect the 8993T>G mutation site. The nondiscriminative (ND) and discriminative (D) assays were mixed and measured with Rotor-Gene Multiplex PCR Kit (Qiagen). All reactions were run in duplicate with two different amounts of input DNA: 1–4 ng and 1:8 dilutions. The SDS software generated a standard curve using four eightfold dilutions plus a final fourfold dilution. The percentage of mtDNA carryover in relation to the total mtDNA content was calculated by the equation: heteroplasmy = $100 \times (\text{quantity D}/\text{quantity ND})$. ARMs-qPCR was also applied to detect 8993T>G, 3243A>G and 13513G>A heteroplasmy levels in fibroblasts and iPS cells using primers and TaqMan MGB probes specifically targeting to mutation sites.

Whole mtDNA sequencing analysis by MiSeq. Single PCR amplification of entire human mtDNA was performed with primers mtDNA-F-2120, (GGAGC ACTAGGAAAAACCTTGTAAGAGAG) and mtDNA-R-2119 (AAAGAGC TGTTCCCTCTTTGGACTAACA) under the following conditions: 94 °C for 1 min followed 98 °C for 10 s and 68 °C for 16 min \times 30 cycles and then 72 °C for 10 min. PCR amplifications were performed using TAKARA LA Taq polymerase (Takara Biotechnology) and the concentrations of PCR products were measured using a Qubit 2.0 Fluorometer. The Nextera XT DNA sample preparation kit (Illumina) was used to prepare the libraries. Sequencing was performed on an Illumina MiSeq instrument and the data were analysed using NextGENe software. Briefly, sequence reads ranging from 100 to 200 bp were quality filtered and processed using BLAT algorithm. Sequence error correction feature (condensation) was performed to reduce false-positive variants and produce sample consensus sequence and variant calls. Alignment without sequence condensation was used to calculate percentage of mitochondrial genome with depth of coverage of 1,000. Starting from quality FASTQ reads, the reads were quality filtered and converted to FASTA format. Filtered reads were then aligned to the revised Cambridge Reference Sequence (rCRS) of the human mtDNA (NC_012920.1) followed by variant calling. Variant heteroplasmy was calculated by NextGENe software as follows: Base heteroplasmy (mutant allele frequency %) = mutant allele (forward + reverse)/total coverage of all alleles C, G, T, A (forward + reverse) \times 100. The clinical significance of the variants was then analysed with MitoMaster (<http://www.mitomap.org/MITOMASTER/WebHome>).

Live cell oxygen consumption. XF24 or XF96 extracellular flux analysers (Seahorse Biosciences) were used to measure oxygen consumption rates (OCR) as described¹². In brief, stem-cell-derived fibroblasts were seeded at a density of 50,000 cells per well of a XF24 cell culture microplate and incubated for 24 h to ensure attachment. Before assay, cells were equilibrated for 1 h in unbuffered XF assay medium supplemented with 25 mM glucose, 1 mM sodium pyruvate, 2 mM glutamax, 1 \times nonessential amino acids and 1% FBS in a non-CO₂ incubator. Mitochondrial processes were examined through sequential injections of oligomycin (0.5 μ g ml⁻¹), carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP, 1 μ M) and rotenone (0.5 μ M)/antimycin A (1 μ M). Indices of mitochondrial function were calculated as basal respiration rate (baseline OCR – rotenone/antimycin A OCR), ATP dependent (basal respiration rate – oligomycin OCR), maximal respiration rate (FCCP OCR – rotenone/antimycin A OCR) and oxidative reserve (maximal respiration rate – basal respiration rate). For other cell types, an XF96 extracellular flux analyser was used with 20,000 cells seeded to each well of a XF96 cell culture microplate. After a 24-h attachment period, mitochondrial processes were examined using the same protocol as above. Each plotted value was normalized to total protein quantified using a Bradford protein assay (Bio-rad). Results were presented as mean \pm s.e.m. One-way ANOVA was used for three group comparisons and Student's *t*-test was used for two group comparisons. A *P* value less than 0.05 was considered significant.

Flow cytometric analysis. The efficiency of differentiation protocols was assessed by FACS. For cardiomyocyte differentiation, 0.25 million cells were fixed in the presence of 1% (vol/vol) paraformaldehyde at room temperature for 20 min. Fixed cells were then incubated in 90% (vol/vol) cold methanol for 15 min at 4 °C, rinsed

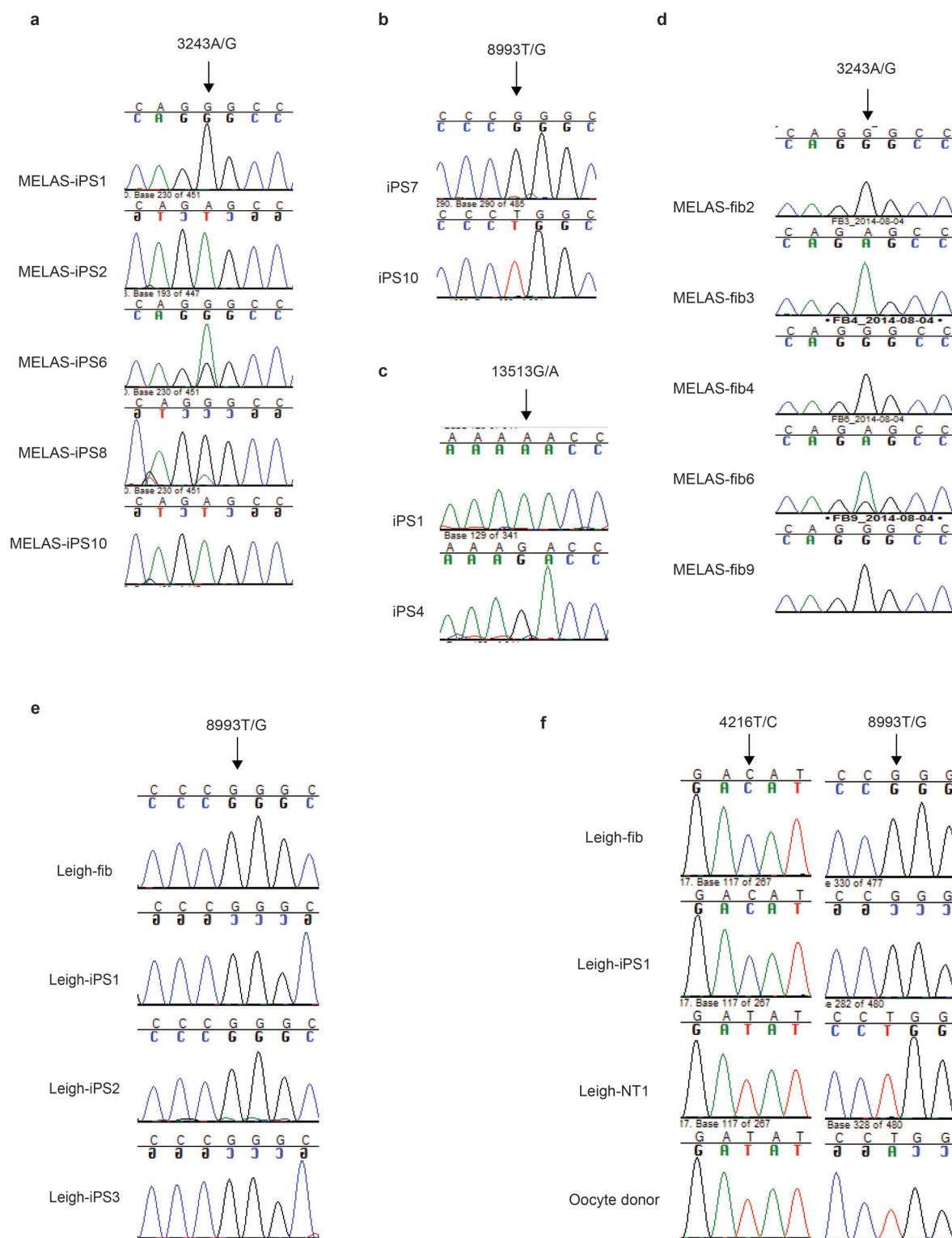
two times and incubated overnight at 4 °C with a primary antibodies against GATA4 (Santa Cruz) and cTnT (Pierce). After staining, cells were rinsed two times and incubated in the presence of 1:1,000 secondary antibodies (donkey Alexa 488 and 567; Molecular Probes) for 30 min. After staining, cells were washed two times and re-suspended for analysis.

Real time RT-PCR. RNA was isolated using RNeasy kit (Qiagen) as per manufacturer's instructions. cDNA synthesis was performed using the iScript™ cDNA synthesis kit for RT-PCR (BioRad). Real-time PCR was performed using the SYBR Green Supermix (BioRad). The levels of expression of respective genes were normalized to corresponding 18S values and are shown as fold change relative to the value of the control sample. All reactions were done in triplicate.

RNA-seq library construction and data analysis. RNA was isolated with Micro-to-Midi Total RNA Purification System (Life Technologies), quality evaluated (RNA6000 Nano Kit and BioAnalyzer 2100, Agilent) made into sequencing libraries, sequenced and mapped as previously described²⁶. Libraries were constructed using 500 ng input RNA per sample. Approximately 27 million reads were generated per sample, and 73% of these reads were uniquely mapped. Counts for each gene were quantified using the python script `rpkmforgenes` and annotated using Ensembl GRCh37. Genes without at least one sample with at least five reads

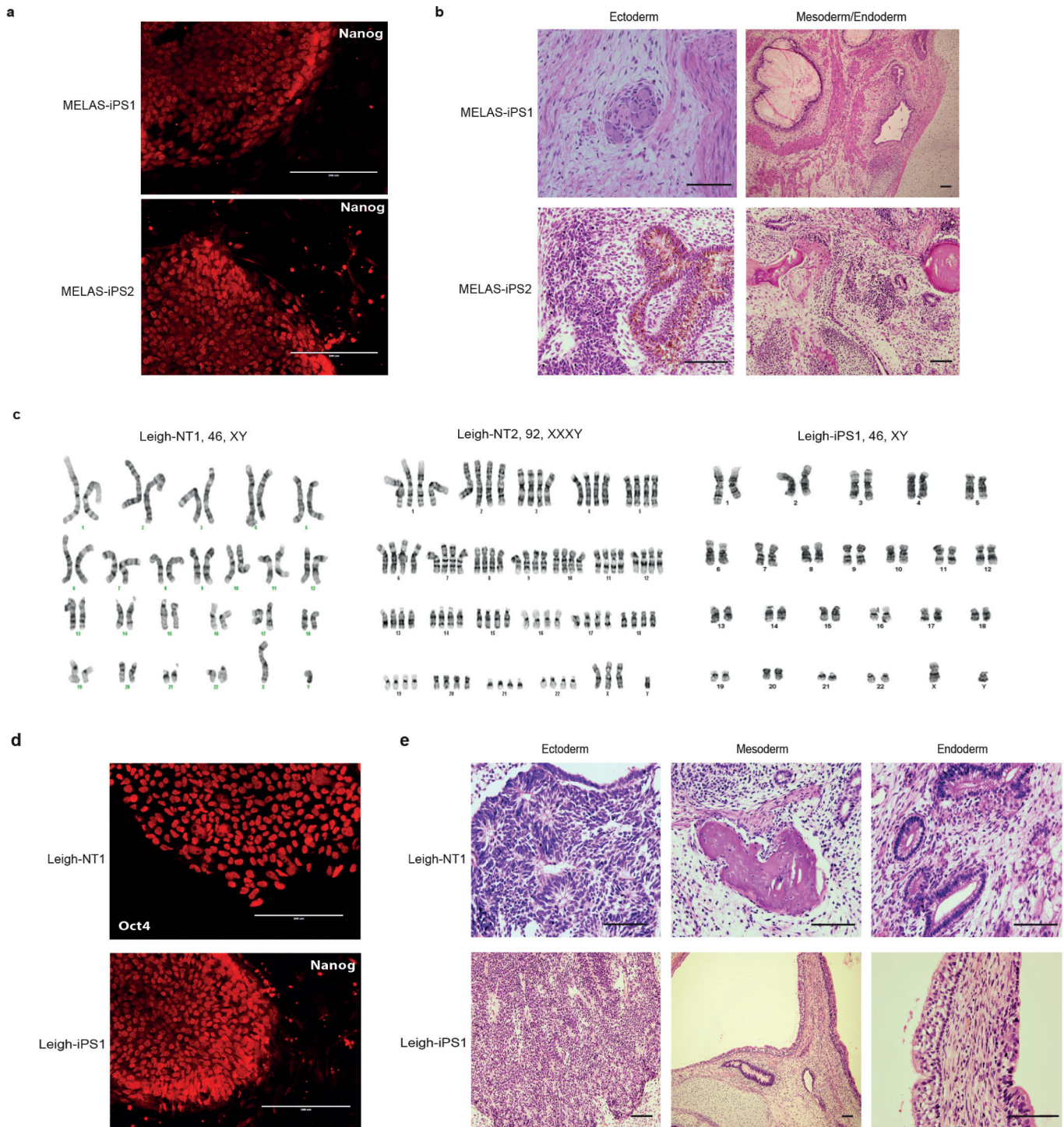
were removed from the analysis. The count data was normalized and differential expression was performed using the R (v.3.1.1) package DESeq2 (v.1.4.5). Briefly, DESeq2 uses negative binomial generalized linear models and shrinkage estimation for dispersions and fold changes to improve stability and interpretability of the estimates²⁹. It reports a *P* value and an adjusted *P* value using the Benjamini–Hochberg procedure. Genes with an adjusted *P* value less than 0.05 were considered differentially expressed unless otherwise noted. Heat maps were constructed using the R (v.3.1.1) package `gplots` (v.2.14.2). Each variable was standardized by subtraction of its mean value and division by its standard deviation across all samples. All functional enrichment analyses were generated using the Genomic Regions Enrichment of Annotations Tool (v. 2.0.2)³⁰ with default settings. Hierarchical clustering was performed with the R package `pvcust`, with Euclidian distance and average linkage with 10,000 bootstraps.

28. Paull, D. *et al.* Nuclear genome transfer in human oocytes eliminates mitochondrial DNA variants. *Nature* **493**, 632–637 (2013).
29. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).
30. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnol.* **28**, 495–501 (2010).



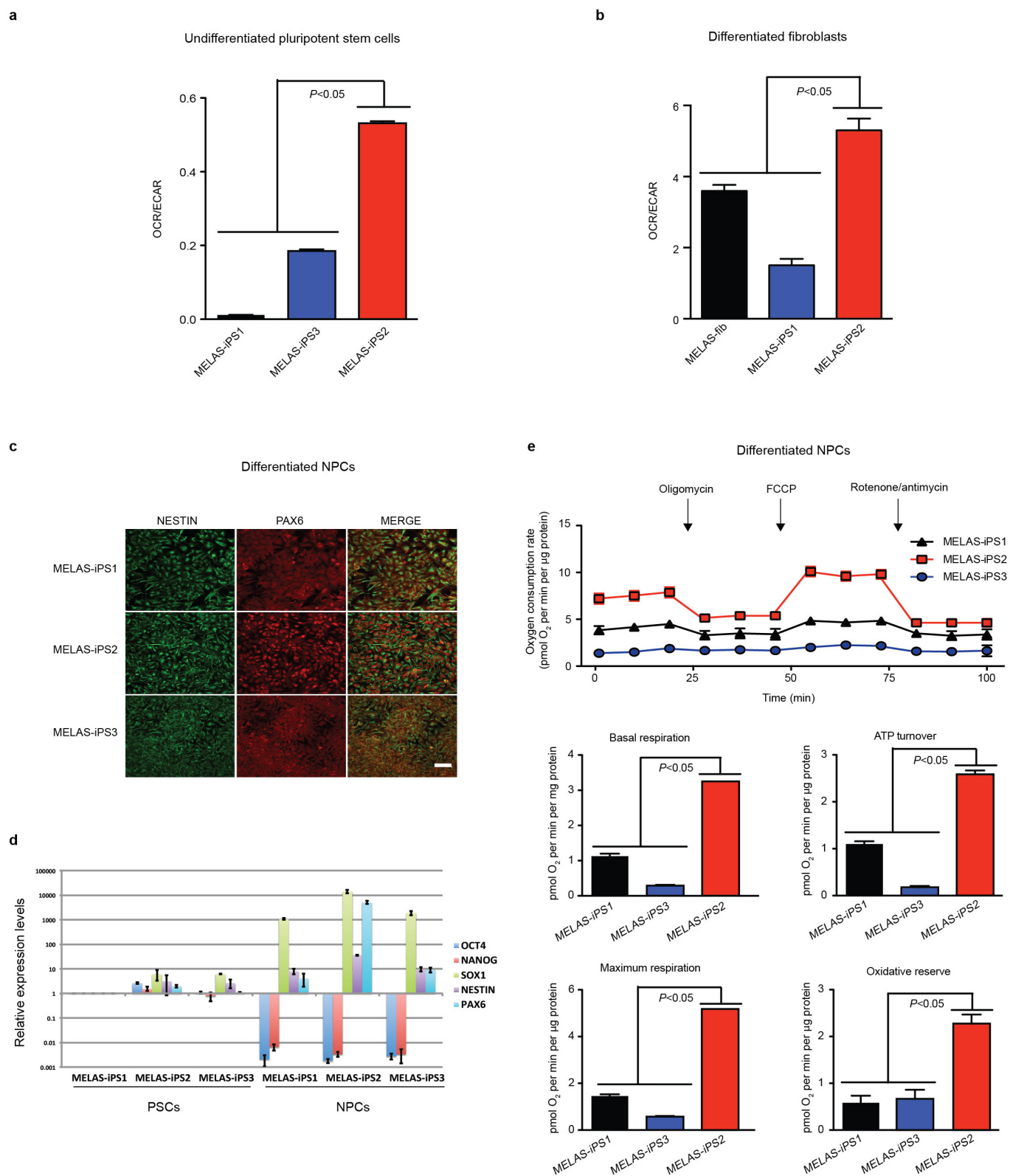
Extended Data Figure 1 | mtDNA genotyping by Sanger sequencing.
a, Chromatographs showing mtDNA genotyping at 3243 position (arrow) in representative MELAS iPS cells. **b**, Chromatographs showing mtDNA genotyping at 8993 position (arrow). **c**, mtDNA at 13513 position (arrow) in representative iPS cells derived from Leigh syndrome patients. **d**, Chromatographs showing either wild-type A or mutant G allele at position 3243 in

representative MELAS fibroblast clones. **e**, mtDNA genotyping demonstrated that all Leigh-iPS cell lines and Leigh-fib contain a G mutation allele at mtDNA position 8993. **f**, mtDNA genotyping demonstrated that Leigh-fib and Leigh-iPS1 cell lines contained a C mutant allele at position 4216 and a G mutant allele at position 8993, while Leigh-NT1 line carried oocyte mtDNA with a wild-type T allele at both positions.



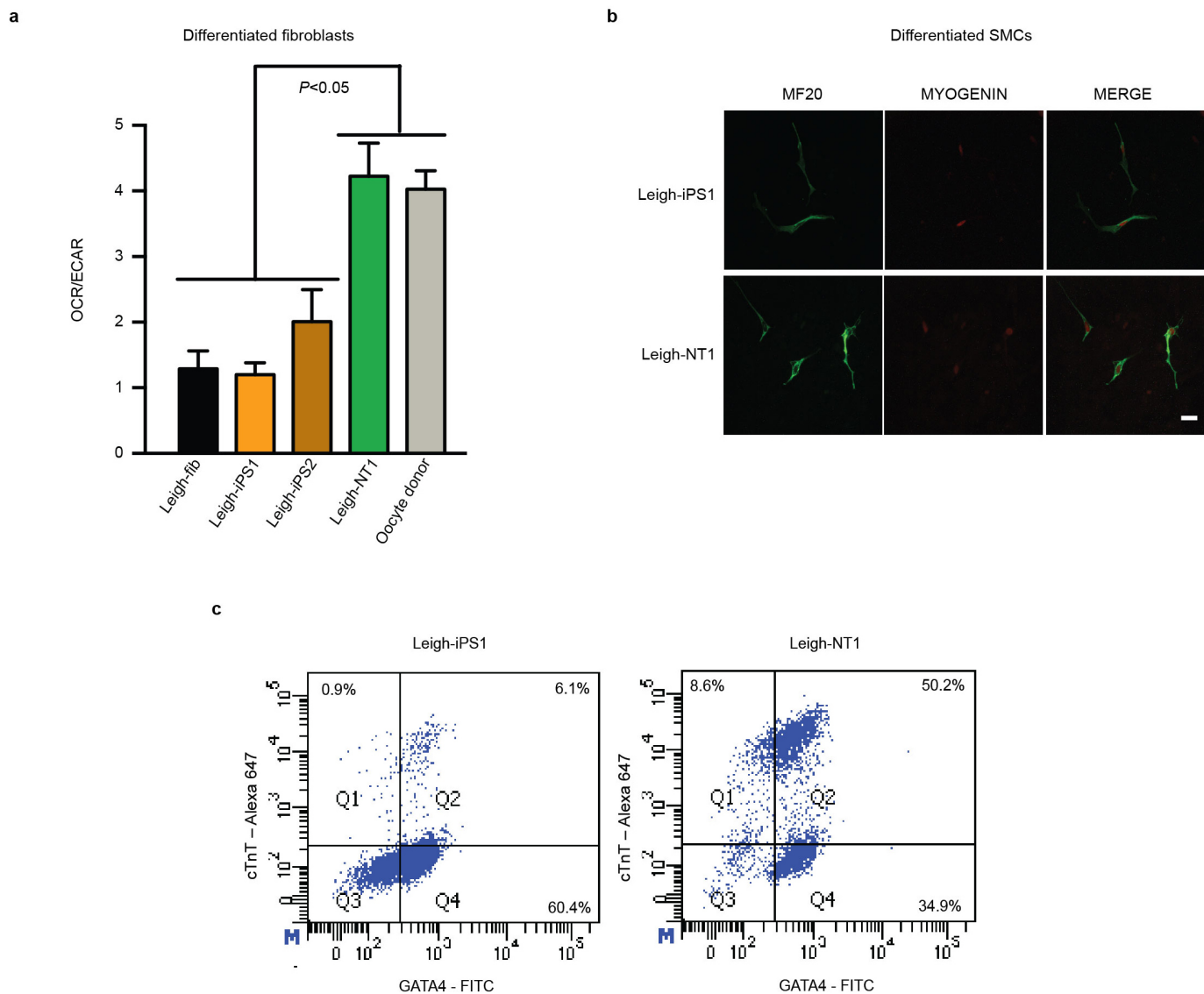
Extended Data Figure 2 | Cytogenetic, pluripotency and teratoma analyses. **a**, MELAS-iPS1 and MELAS-iPS2 expressing NANOG detected by immunocytochemistry. Scale bars, 200 μ m. **b**, Histological analyses of teratoma tumours produced after injections of MELAS-iPS1 and MELAS-iPS2 cells into SCID mice. Scale bars, 200 μ m. **c**, Cytogenetic G-banding analysis confirmed that Leigh-NT1 and Leigh-iPS1 exhibited normal 46XY karyotypes and Leigh-NT2

exhibited a XXXY tetraploid karyotype. **d**, Leigh-NT1 and Leigh-iPS1 cells expressed OCT4 and NANOG. Scale bars, 200 μ m. **e**, Histological analyses of teratoma tumours produced after injections of Leigh-NT1 and Leigh-iPS1 cells into SCID mice. Scale bars, 200 μ m. Haematoxylin and eosin staining of teratoma sections identify derivatives of ectoderm, mesoderm and endoderm.



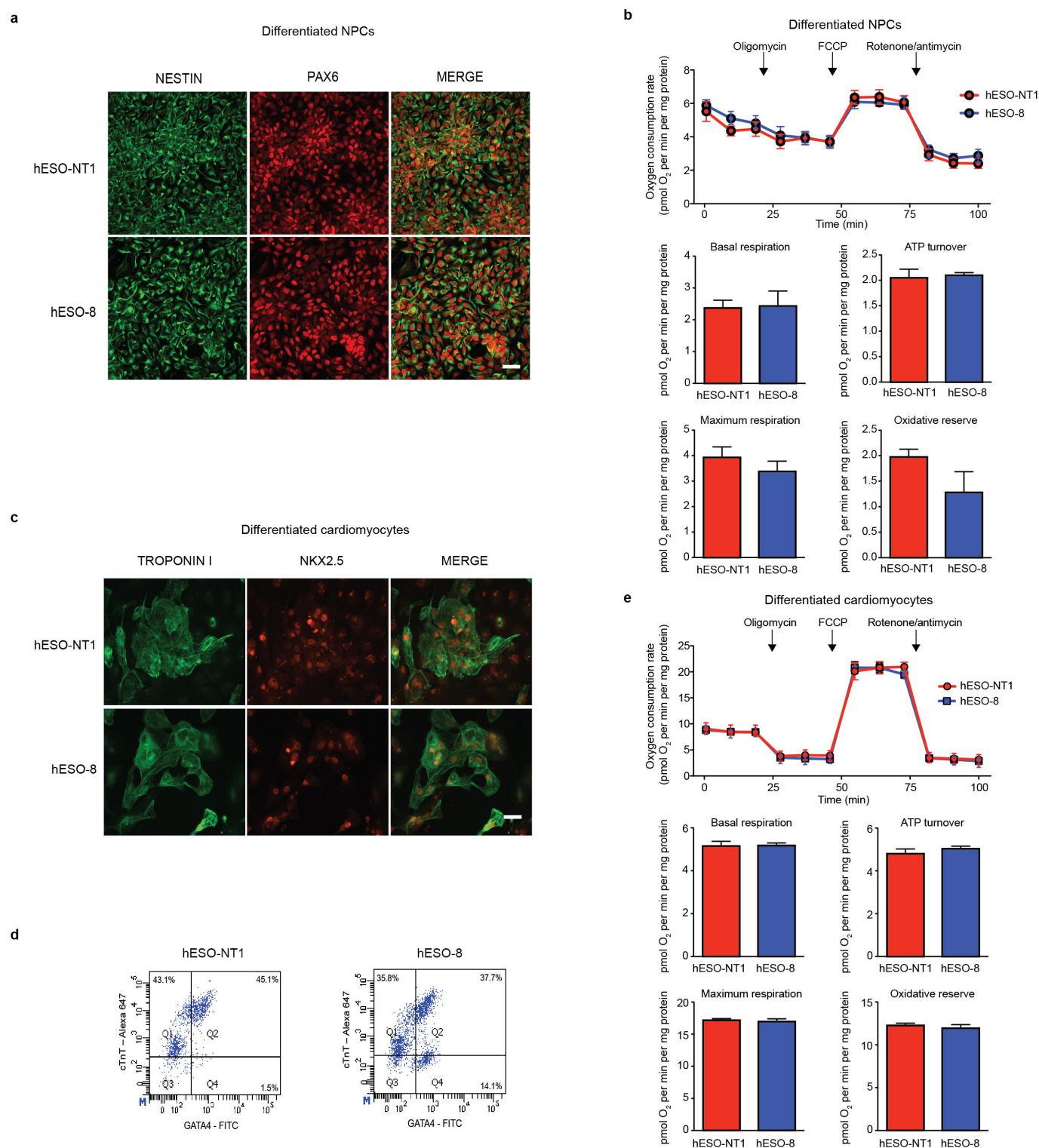
Extended Data Figure 3 | Metabolic function in differentiated cells from MELAS iPS cells. **a**, OCR/ECAR ratio in MELAS-iPS cells. Mutant MELAS-iPS1 and MELAS-iPS3 displayed significantly decreased OCR/ECAR ratios compared to wild-type MELAS-iPS2 ($P < 0.05$), indicating a greater reliance on glycolysis ($n = 9$ per cell line, biological replicates). **b**, OCR/ECAR ratio in MELAS-iPS cell derived fibroblasts ($n = 10$ per cell line, biological replicates). **c**, Immunofluorescence analysis for neural progenitor markers in MELAS-iPS

derived NPCs. Scale bar, 100 μm . **d**, Quantitative analysis of PSC (OCT4 and NANOG) or NPC (SOX1, NESTIN and PAX6) marker expression in MELAS-iPS cells and NPCs ($n = 3$ per cell line, biological replicates). **e**, OCR of MELAS-iPS cell derived NPCs ($n = 6$ per cell line, biological replicates). Error bars are mean \pm s.e.m. Significance established with one-way ANOVA with Tukey's multiple comparison test.



Extended Data Figure 4 | Metabolic function in differentiated cells from Leigh syndrome PSCs. **a**, OCR/ECAR ratio in Leigh-iPS1, Leigh-iPS2 and Leigh-NT1 derived fibroblasts, parental and oocyte donor fibroblasts ($n = 9, 8, 10, 9$ and 8 per cell line, respectively, biological replicates). **b**, Immunofluorescence analysis of Leigh-iPS1- and Leigh-NT1-derived skeletal muscle cells labelled with MF20 and myogenin antibodies. Scale bar, $100 \mu\text{m}$.

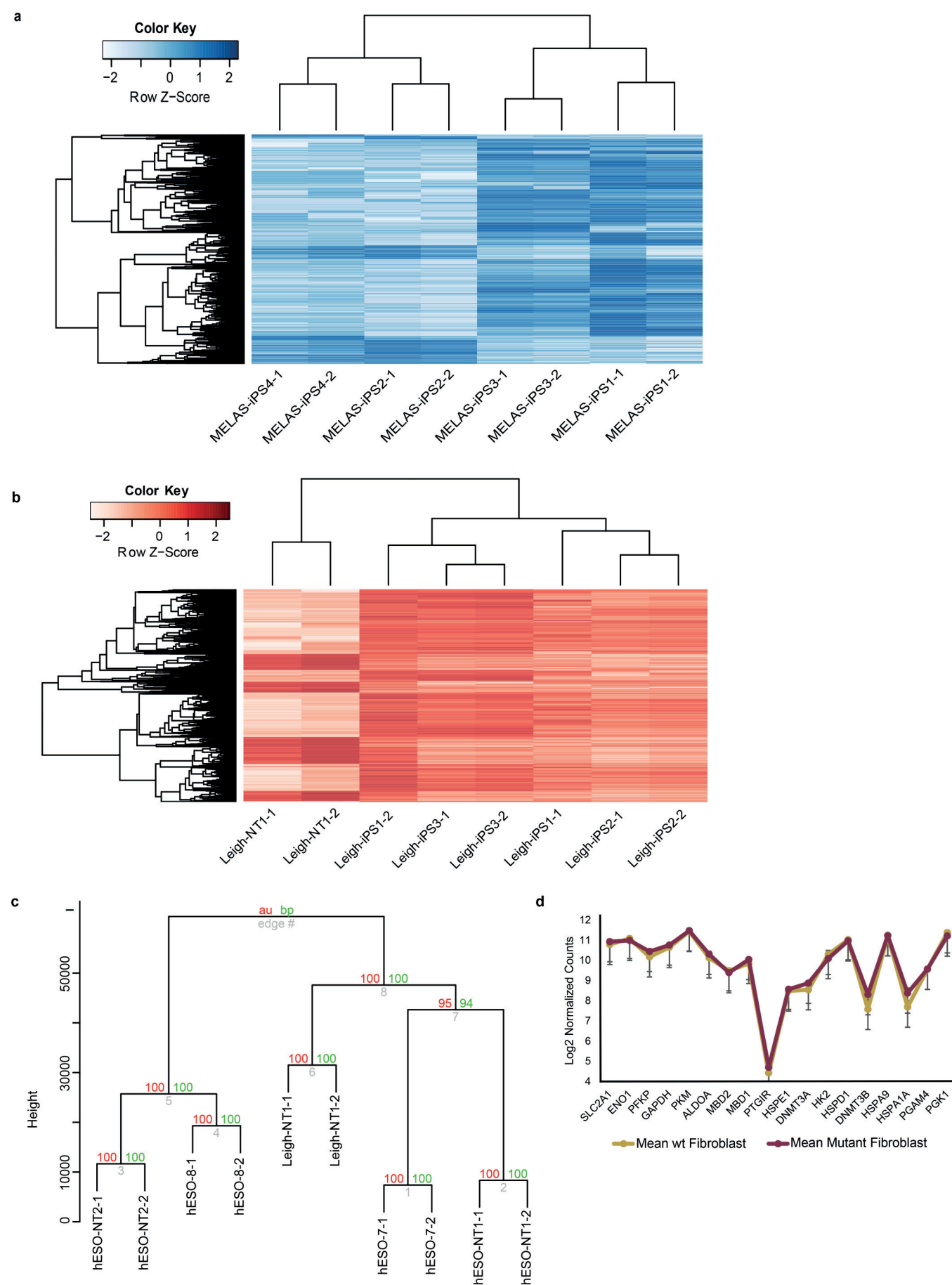
c, Cardiomyocyte differentiation efficiency in Leigh-iPS1 and Leigh-NT1 evaluated by FACS for CTnT-Alexa 647 and GATA4-FITC antibodies ($n = 3$ per cell line, biological replicates). Error bars are mean \pm s.e.m. Significance established with one-way ANOVA with Tukey's multiple comparison test.



Extended Data Figure 5 | Metabolic function in hESO-NT1 and hESO-8.

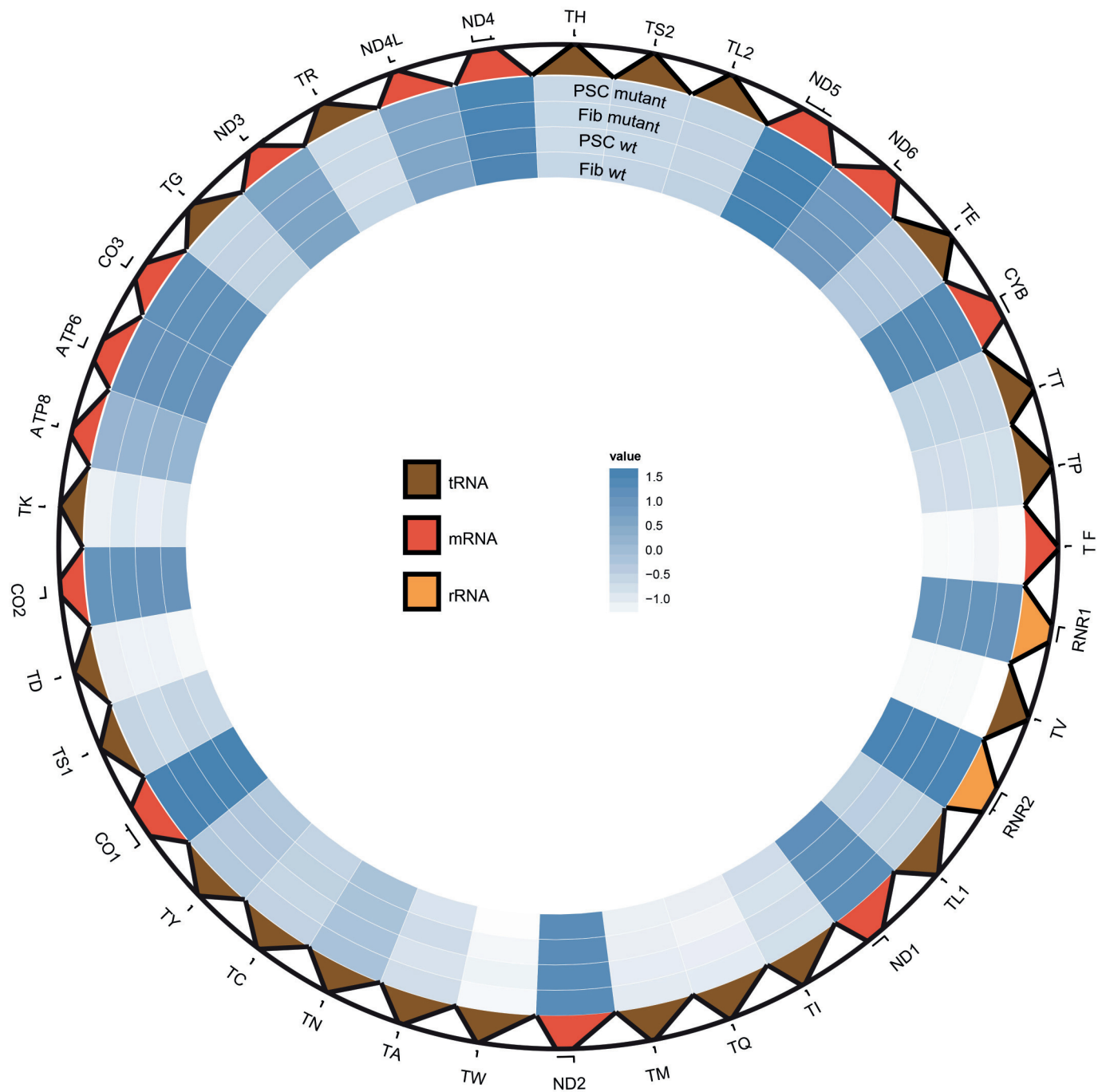
a, Immunofluorescence analysis of hESO-NT1 and hESO-8 derived NPCs with nestin and PAX6 antibodies. Scale bar, 100 μ m. **b**, Metabolic profiles of NPCs differentiated from hESO-NT1 and hESO-8 ($n = 6$ per cell line, biological replicates). **c**, Immunofluorescence analysis of hESO-NT1 and hESO-8 derived cardiomyocytes with troponin I and NKX2.5 antibodies. Scale bar, 100 μ m.

d, Efficiency of cardiomyocyte differentiation in hESO-NT1 and hESO-8 evaluated by FACS analysis for cTnT-Alexa 647 and GATA4-FITC antibodies ($n = 3$ per cell line, biological replicates). **e**, OCR of hESO-NT1 and hESO-8 derived cardiomyocytes ($n = 6$ per cell line, biological replicates). Error bars are mean \pm s.e.m. Significance established with Student's t -test.



Extended Data Figure 6 | RNA-seq analyses of fibroblasts differentiated from MELAS and Leigh syndrome PSCs carrying wild-type and mutant mtDNA. **a**, Heat map showing all differentially expressed 1,118 genes (adjusted P value < 0.05) between fibroblasts differentiated from mutant MELAS iPS cells ($n = 4$ from biological duplicates of MELAS-iPS2 and MELAS-iPS4) and wild-type MELAS iPS cells ($n = 4$ from biological duplicates of MELAS-iPS1 and MELAS-iPS3). **b**, Heat map demonstrating differentially expressed 2,950 genes (adjusted P value < 0.05) between fibroblasts derived from wild-type Leigh-NT1 (biological duplicates) and mutant Leigh iPS cells ($n = 6$ from biological duplicates of Leigh-iPS1, Leigh-iPS2 and Leigh-iPS3). **c**, Hierarchical clustering using Euclidean distance and average linkage using

pvclust, which employs a multiple bootstrap resampling algorithm to calculate the approximately unbiased (AU, red) and bootstrap probability (BU, green) values for cluster distinctions. Hierarchical clustering showed that the Leigh-NT1 fibroblasts were similar to hESO-NT1, hESO-NT2, hESO-7 and hESO-8 fibroblasts. **d**, Mean \log_2 normalized counts \pm s.e.m. for genes previously reported to be differentially expressed in MELAS cytoplasmic hybrid clones and involved in metabolic and stress response, signalling pathways and epigenetic modifying processes (wild type fibroblast; $n = 14$ from biological duplicates of 7 independent cell lines; mutant fibroblast $n = 10$ from biological duplicates of 5 independent cell lines).



Extended Data Figure 7 | RNA-seq analysis of the mitochondrial transcriptome. Circular heat map displaying average expression levels for all mitochondrial genes grouped by sample differentiation status and presence or absence of a mutation in the mitochondrial genome (Fib mutant (including primary fibroblasts and PSC derived fibroblasts) with mutant mtDNA $n = 14$, biological duplicates of 7 independent cell lines; Fib wild type (including

primary fibroblasts and PSC derived fibroblasts) with wild-type mtDNA $n = 14$, biological duplicates of 7 independent cell lines; PSC mutant (undifferentiated IVF-ESC, NT-ESC and iPS cells) with mutant mtDNA $n = 3$; PSC wild type (undifferentiated IVF-ESC, NT-ESC and iPS cells) with wild-type mtDNA $n = 12$). The expression of mtDNA-encoded genes was similar irrespective of 3243A>G or 8993T>G mutations (adjusted P value >0.05).

Extended Data Table 1 | Mutation loads in Leigh syndrome iPS cells with homoplasmic mutations

PSC lines	8993 T>G mutant G %	4216 T>C mutant C %
Leigh-iPS1	100	100
Leigh-iPS2	100	100
Leigh-iPS3	100	100
Leigh-iPS4	100	100
Leigh-iPS5	100	100
Leigh-NT1	0	0

Extended Data Table 2 | Quantitative mutant mtDNA carryover analysis in Leigh-NT1

Cell line	% mutant mtDNA P5 (±SD)	% mutant mtDNA P15 (±SD)	% mutant mtDNA P20 (±SD)	% mutant mtDNA P30 (±SD)	% mutant mtDNA P40 (±SD)
Leigh-NT1	0.14±0.06	undetectable	undetectable	undetectable	undetectable

Extended Data Table 3 | Short tandem repeat analysis of oocyte donors, Leigh-NT2 and iPS cells from the Leigh syndrome patient

Sample	Leigh group					
	Leigh-fib	Leigh-iPS1	Leigh-NT1	Oocyte donor1	Leigh-NT2	Oocyte donor2
Sex AME	M XY	M XY	M XY	F XX	F+M XXXY	F XX
D1S548	152/172	152/172	152/172	152/168	152/168/172/176	168/176
D2S1333	301/305	301/305	301/305	285/301	285/301/305	285/301
D3S1768	188/200	188/200	188/200	188/196	188/192/200	192/192
D4S2365	288/296	288/296	288/296	296/300	280/288/296	280/296
D4S413	123/123	123/123	123/123	123/123	123/155/157	155/157
D5S1457	119/123	119/123	119/123	123/123	119/123	119/123
D6S501	168/172	168/172	168/172	168/172	168/172/176	172/176
D7S513	179/197	179/197	179/197	187/201	179/187/189/197	187/189
D9S921	183/183	183/183	183/183	183/189	183/183	183/183
D10S1412	156/162	156/162	156/162	159/162	156/162/165	162/165
D11S2002	254/254	254/254	254/254	254/254	254/254	254/254
D11S925	303/305	303/305	303/305	303/305	282/295/303/305	282/295
D12S364	266/270	266/270	266/270	270/272	266/270/278	270/278
D12S67	260/268	260/268	260/268	252/260	252/256/260/268	252/256
D13S765	192/196	192/196	192/196	188/200	188/192/196	188/192
D16S403	137/141	137/141	137/141	145/145	135/137/141	135/141
D17S1300	257/273	257/273	257/273	257/269	257/265/269/273	265/269
D18S537	196/200	196/200	196/200	196/208	196/200/204	200/204
D18S72	305/305	305/305	305/305	301/305	305/305	305/305
DXS2506*	286*	286*	286*	282/286	282/286	282/282
D6S291	199/201	199/201	199/201	199/199	199/201	199/201
D6S276	239/239	239/239	239/239	227/235	227/239/251	227/251

* Male samples only show one mark of DXS2506 on the X chromosome.

Live imaging RNAi screen reveals genes essential for meiosis in mammalian oocytes

Sybille Pfender^{1*}, Vitaliy Kuznetsov^{1*}, Michal Pasternak^{1*}, Thomas Tischer¹, Balaji Santhanam¹ & Melina Schuh¹

During fertilization, an egg and a sperm fuse to form a new embryo. Eggs develop from oocytes in a process called meiosis. Meiosis in human oocytes is highly error-prone^{1,2}, and defective eggs are the leading cause of pregnancy loss and several genetic disorders such as Down's syndrome^{3–5}. Which genes safeguard accurate progression through meiosis is largely unclear. Here we develop high-content phenotypic screening methods for the systematic identification of mammalian meiotic genes. We targeted 774 genes by RNA interference within follicle-enclosed mouse oocytes to block protein expression from an early stage of oocyte development onwards. We then analysed the function of several genes simultaneously by high-resolution imaging of chromosomes and microtubules in live oocytes and scored each oocyte quantitatively for 50 phenotypes, generating a comprehensive resource of meiotic gene function. The screen generated an unprecedented annotated data set of meiotic progression in 2,241 mammalian oocytes, which allowed us to analyse systematically which defects are linked to abnormal chromosome segregation during meiosis, identifying progression into anaphase with misaligned chromosomes as well as defects in spindle organization as risk factors. This study demonstrates how high-content screens can be performed in oocytes, and allows systematic studies of meiosis in mammals.

Meiosis is still much more poorly understood than mitosis, especially in mammals. Systematic screens have greatly increased our understanding of mitosis. However, high-content screens for

mammalian meiotic genes have so far been precluded by various technical challenges. For instance, mammalian oocytes are only available in small numbers; genetic screens in mammals are slow; and RNA interference (RNAi) in oocytes is inefficient owing to large amounts of stored protein. Oocytes accumulate proteins while they grow within follicles in the ovary⁶. Thus, we established a protocol that allowed us to block protein expression by RNAi during follicle growth and subsequently to assess gene function by quantitative live imaging (Fig. 1a). Briefly, we microinjected short interfering RNAs (siRNAs) into small follicle-enclosed oocytes and grew the follicles *in vitro*, combining and modifying previous methods^{7–9}. When the oocytes had reached their full size, we isolated and labelled them, and imaged meiosis live for around 18 h on confocal microscopes using automated imaging routines.

The oocytes grown *in vitro* resembled those grown *in vivo*: first, the efficiency of nuclear envelope breakdown (NEBD) and polar body extrusion, as well as the timing of meiotic progression, were similar (Fig. 1b–d and Extended Data Fig. 1d, e); second, their transcriptome was related (Extended Data Fig. 2a–c and Supplementary Table 1); third, they developed into blastocysts with similar efficiency upon fertilization (Extended Data Fig. 1f, g)⁸.

Follicle culture and microinjection are labour-intensive, precluding genome-wide screens. Instead, we preselected 774 target genes that were highly expressed in mouse oocytes, while excluding messenger RNAs (mRNAs) stored for embryo development. We took advantage

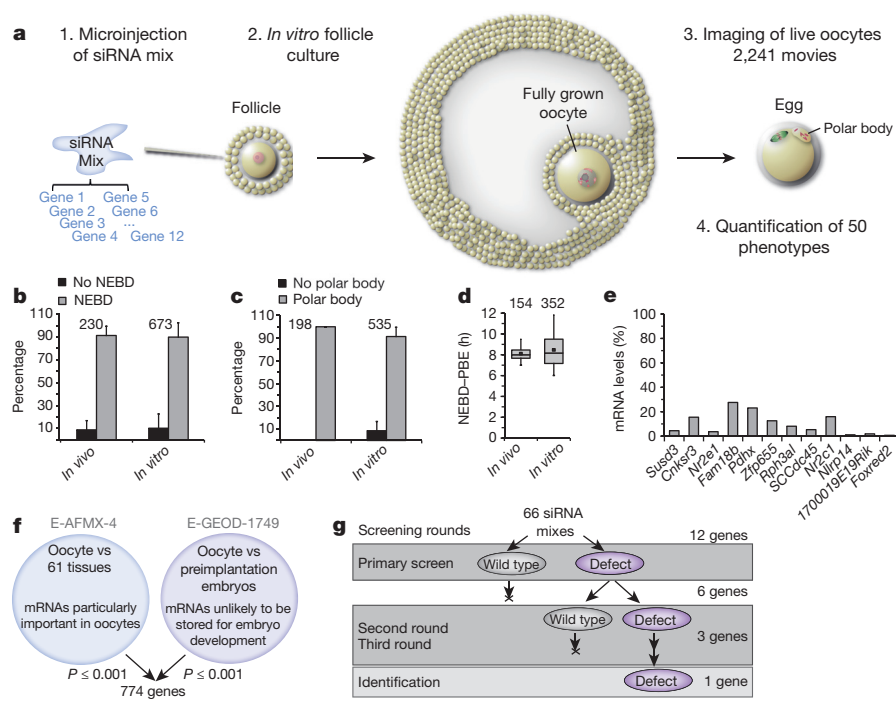


Figure 1 | RNAi screen in live oocytes.

a, Principle of the screen. **b–d**, Meiotic progression in oocytes grown *in vivo* (5 independent experiments) and *in vitro* (42 independent experiments). The box plot in **d** shows median (line), mean (small square), 5th, 95th (whiskers) and 25th and 75th percentile (boxes). Number of oocytes is given next to bars. PBE, polar body extrusion. **e**, mRNA depletion upon targeting 12 genes simultaneously by RNAi. Representative example for four independent experiments. Errors bars in **b–d**, s.d. **f**, Selection of target genes. **g**, Strategy to identify genes in mixes causing phenotypes.

¹Medical Research Council, Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK.

*These authors contributed equally to this work.

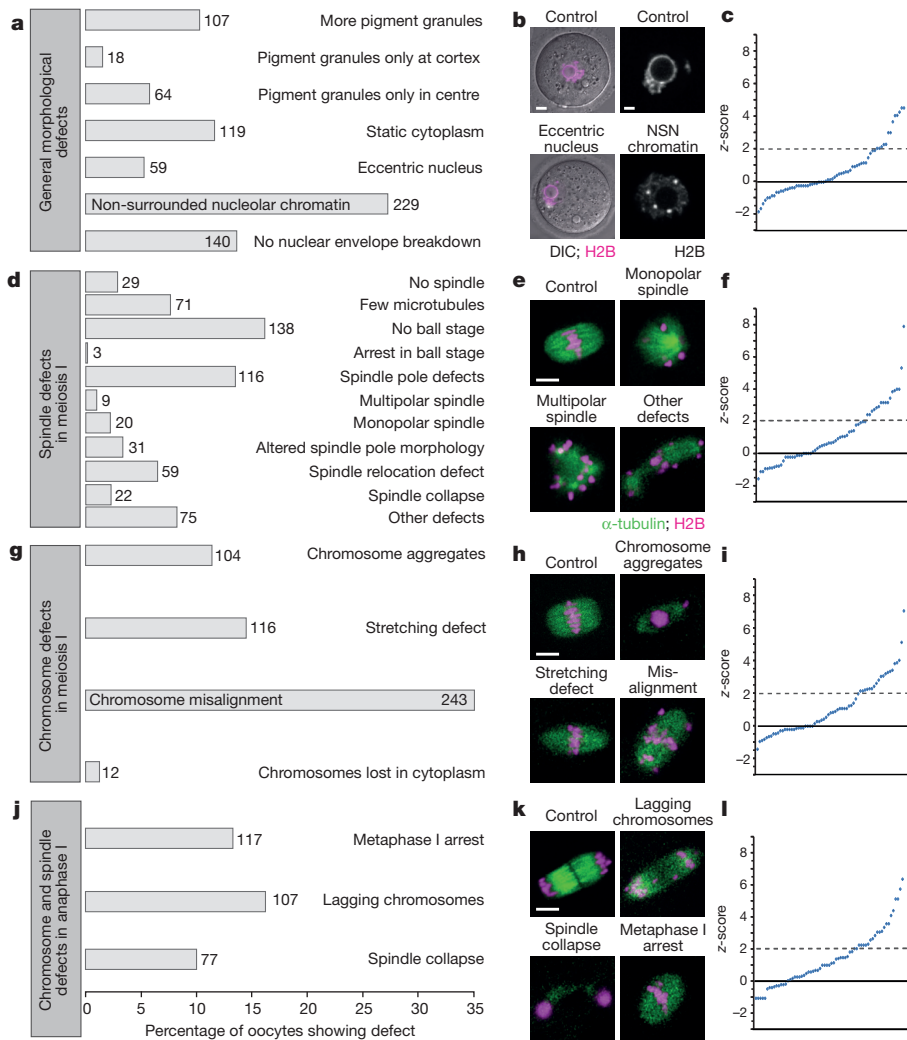


Figure 2 | Defects during meiosis I in siRNA-treated oocytes. **a, d, g, j.** Frequency of defects in siRNA-treated oocytes. Number of oocytes is given next to bars. Seventy independent experiments. **b, e, h, k.** Representative examples of phenotypes quantified in **a, d, g, j.** Chromosomes (magenta; white in **b**), microtubules (green). NSN, non-surrounded nucleolus chromatin configuration. Scale bar, 10 μ m. **c, f, i, l.** siRNA mixes sorted according to z-score in each category. Dashed line delineates mixes with a z-score higher than two s.d. above the average value of all controls.

of two microarray data sets^{10,11}, which compare the expression profile of oocytes with the profiles of other cell types and preimplantation mouse embryos, respectively (Fig. 1f). Only genes that were significantly upregulated in oocytes in both data sets were selected for the screen.

To achieve high throughput, we targeted 12 genes simultaneously (Fig. 1a, e). Co-depletion of several genes led to the expected phenotype for genes with known functions. For instance, mixes targeting one of the zona pellucida genes (*Zp1*, *Zp2* or *Zp3*) together with 11 other genes prevented formation of the zona¹² (Extended Data Fig. 1h, i), and mixes targeting spindle assembly checkpoint proteins led to the expected earlier onset of anaphase (for example, mix 33P1-2-3-4-5-6 targeting *Bub1* in Supplementary Table 2).

The targeting of all 774 genes enabled us to make videos of 2,241 individual oocytes, including 1,210 RNAi-treated and 1,031 control oocytes. We scored every oocyte for 41 possible defects and determined 5 characteristic meiotic time points as well as the spindle length and width in meioses I and II (Supplementary Table 3; scored parameters are described in Extended Data Fig. 3). The frequencies with which different defects were observed in RNAi-treated and control oocytes are plotted in Fig. 2 and Extended Data Figs 4a–i and 5. To identify significant hits, we calculated the z-score of individual mixes for different categories. These quantifications resulted in a comprehensive annotated resource of defects (Supplementary Tables 2 and 3 and Extended Data Fig. 2d). Supplementary Table 2 allows users to easily query if siRNA mixes targeting their gene of interest result in

defects in oocyte meiosis or to identify mixes causing defects in the stage of meiosis they are studying. This Supplementary Table also includes hyperlinks to the original video files for further assessment.

Proof-of-principle experiments demonstrated that defects observed upon targeting several genes simultaneously could be allocated to individual genes by stepwise splitting of siRNAs into smaller pools (Fig. 1g and Extended Data Fig. 4j). Several of the identified genes have not yet been implicated in mouse oocyte meiosis, demonstrating that this screening strategy is suitable to identify new meiotic genes. Hits were verified typically three times when the siRNA mixes were split to track down the genes that caused the phenotype of interest. To confirm the observed defects, siRNAs were microinjected again upon gene identification. In addition, specificity was confirmed by microinjection of individual siRNAs and rescue experiments as detailed below.

The screen identified several genes that control meiotic progression, including *Dusp7*, a poorly characterized dual-specificity phosphatase. More than 40% of *Dusp7*-depleted oocytes failed to undergo NEBD (Fig. 3a, b and Supplementary Video 1). In the remaining 60%, NEBD was significantly delayed (Fig. 3c). NEBD could be rescued by wild-type DUSP7 fused with enhanced green fluorescent protein (eGFP-DUSP7), but not by the catalytically inactive eGFP-DUSP7 C333S mutant (Fig. 3b), indicating that the phosphatase activity of DUSP7 is essential for NEBD. eGFP-DUSP7 was excluded from the nucleus (Fig. 3d), suggesting that it promotes NEBD by dephosphorylating cytoplasmic proteins. Together, these data identify *Dusp7* as a phosphatase essential for NEBD in oocytes.

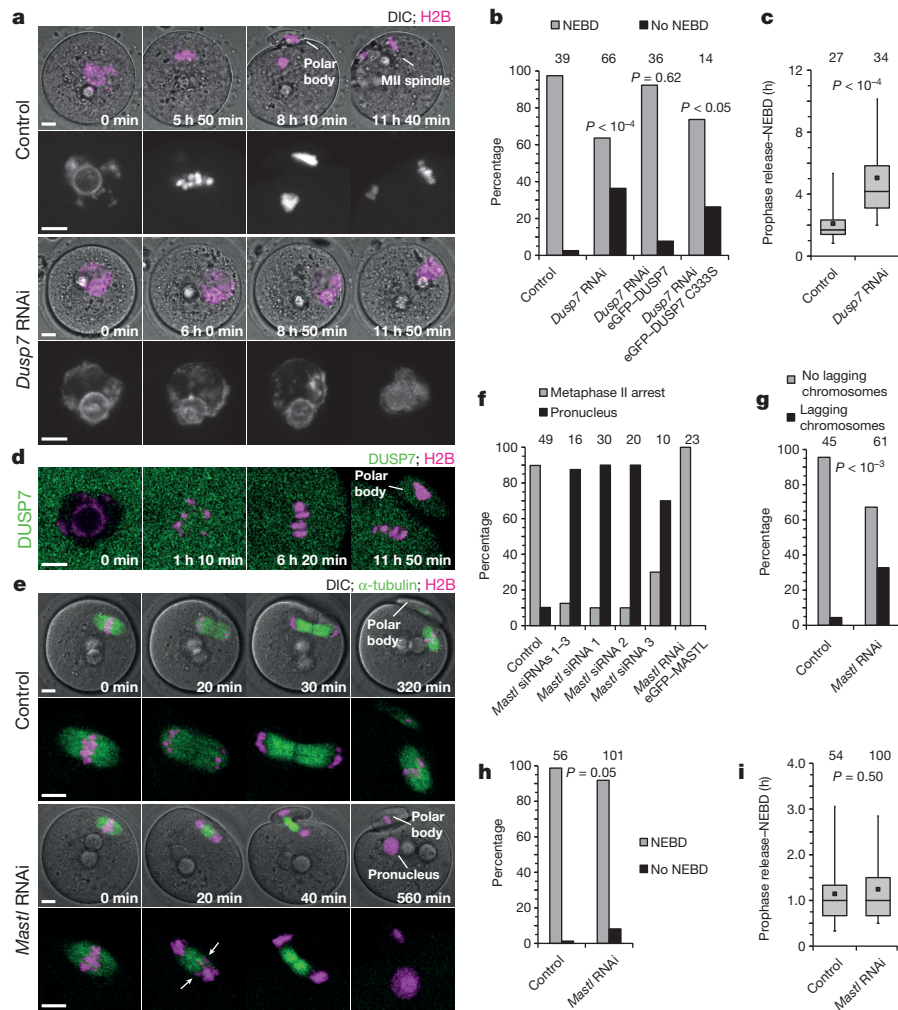


Figure 3 | *Dusp7* and *Mastl* depletion phenotypes. **a**, Oocytes microinjected with control or *Dusp7* siRNAs. Chromosomes in magenta. Quantification of phenotype in **b**, **c**. Scale bar, 10 μ m. **b**, **c**, Efficiency (**b**) and timing of NEBD (**c**) in oocytes microinjected with *Dusp7* siRNAs alone or together with mRNA encoding eGFP-DUSP7 or eGFP-DUSP7 C333S. **d**, Localization of DUSP7 during oocyte maturation. Live oocytes expressing DUSP7 fused with eGFP (green) and H2B fused with monomeric red fluorescent protein (mRFP; magenta, chromosomes). Scale bar, 10 μ m. Representative for 36 oocytes from 5 experiments. **e**, Oocytes microinjected with control or *Mastl* siRNAs. Microtubules in green, chromosomes in magenta. Arrows highlight lagging chromosomes. Quantification of phenotypes in **f-i**. Scale bar, 10 μ m. **f-i**, Oocytes microinjected with different *Mastl* siRNAs alone or together with mRNA encoding human eGFP-MASTL were scored for formation of pronuclei (**f**), lagging chromosomes (**g**), and efficiency (**h**) and timing of NEBD (**i**). Number of oocytes is given next to bars. P values were calculated with Fisher's exact (**b**, **g**, **h**) or Student's t -tests (**c**, **i**). Data from six (**b**, **c**), two (**f**) or five (**g-i**) independent experiments. The box plots in **c** and **i** show median (line), mean (small square), 5th, 95th (whiskers) and 25th and 75th percentile (boxes).

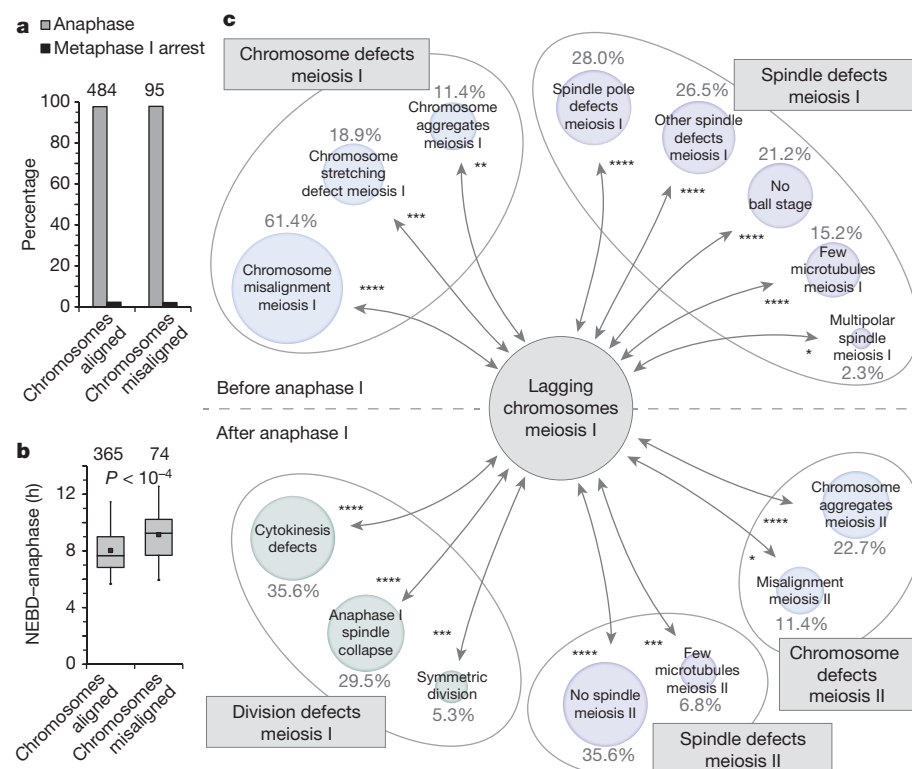


Figure 4 | Factors implicated in chromosome segregation errors. **a**, **b**, The efficiency (**a**) and timing (**b**) of progression into anaphase in control oocytes with aligned and misaligned chromosomes. Number of oocytes is given next to bars. P value was calculated with Fisher's exact test. Data from 52 independent experiments. **c**, Defects significantly more likely to occur in oocytes with lagging chromosomes. Significance was calculated with Fisher's exact test by comparing the prevalence of other defects in oocytes with and without lagging chromosomes, and is specified by asterisks next to arrows, with **** $P < 0.0001$; *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$. The circle area reflects the percentage of oocytes with lagging chromosomes in which each defect was observed.

Another gene essential for meiotic progression was *Eif4enif1*. Mutations in *Eif4enif1* have recently been detected in a family with premature ovarian failure¹³, but the mechanism by which *Eif4enif1* affects fertility is unclear. Our results show that *Eif4enif1* is essential for NEBD and resumption of meiosis (Extended Data Fig. 6a, b).

The screen also provided insights into causes of chromosome segregation errors in oocytes. Several genes were essential for accurate chromosome segregation, including the uncharacterized genes *Fam46b* and *Fam46c* (family with sequence similarity 46), *Aspm*¹⁴ (Extended Data Fig. 7 and Supplementary Video 2), *Birc5* (Survivin)¹⁵ (Extended Data Figs 6), *Ttk*¹⁶ and *Mastl* (Fig. 3e, g). MASTL was also required to prevent exit from meiosis after anaphase I (Fig. 3e, f), but dispensable for meiotic resumption, progression into anaphase, chromosome condensation or cytokinesis (Fig. 3h, i and Extended Data Fig. 8), consistent with a recent study¹⁷.

The screen also allowed us to analyse on a global level how chromosome segregation errors arise in oocytes. With data from 2,241 oocytes, it generated the largest existing data set, to our knowledge, of meiosis in mammalian oocytes (Supplementary Table 2). Evaluation of the control data set identified progression into anaphase with misaligned chromosomes as a major contributor to chromosome segregation errors: misaligned chromosomes only delayed but did not prevent progression into anaphase (Fig. 4a, b). This is consistent with the model that the spindle assembly checkpoint in mammalian oocytes is less stringent than in mitosis^{4,5}.

We were also able to analyse systematically which defects in the oocyte precede chromosomes that lag behind during anaphase. This is of particular interest because lagging chromosomes can lead to inappropriate partitioning of chromosomes upon cytokinesis and are a major cause of aneuploidy^{18,19}. We identified chromosome alignment, individualization and stretching as well as spindle defects as risk factors (Fig. 4c). A systematic representation of how different defects in oocytes were linked is shown in Extended Data Figs 9 and 10.

In summary, we have established an experimental system that now allows systematic studies of meiosis in mammals. The screening approach is scalable and could be adapted to investigate fertilization or embryo development. The follicle-based RNAi method will also be a powerful tool for individual gene studies, as it allows proteins with low turnover to be depleted in oocytes and pre-implantation embryos. The techniques presented in this study should thus facilitate a more rapid accumulation of knowledge about meiosis and early embryo development in mammals, which is crucial to improve methods for treating fertility problems in humans.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 June 2014; accepted 14 May 2015.

Published online 6 July 2015.

1. Kuliev, A., Zlatopolsky, Z., Kirillova, I., Spivakova, J. & Cieslak Janzen, J. Meiosis errors in over 20,000 oocytes studied in the practice of preimplantation aneuploidy testing. *Reprod. Biomed. Online* **22**, 2–8 (2011).
2. Fragouli, E. *et al.* The cytogenetics of polar bodies: insights into female meiosis and the diagnosis of aneuploidy. *Mol. Hum. Reprod.* **17**, 286–295 (2011).

3. Brar, G. A. & Amon, A. Emerging roles for centromeres in meiosis I chromosome segregation. *Nature Rev. Genet.* **9**, 899–910 (2008).
4. Nagaoka, S. I., Hassold, T. J. & Hunt, P. A. Human aneuploidy: mechanisms and new insights into an age-old problem. *Nature Rev. Genet.* **13**, 493–504 (2012).
5. Jones, K. T. & Lane, S. I. Molecular causes of aneuploidy in mammalian eggs. *Development* **140**, 3719–3730 (2013).
6. Clift, D. & Schuh, M. Restarting life: fertilization and the transition from meiosis to mitosis. *Nature Rev. Mol. Cell Biol.* **14**, 549–562 (2013).
7. Jaffe, L. A., Norris, R. P., Freudzon, M., Ratzan, W. J. & Mehlmann, L. M. Microinjection of follicle-enclosed mouse oocytes. *Methods Mol. Biol.* **518**, 157–173 (2009).
8. Eppig, J. J. & Schroeder, A. C. Capacity of mouse oocytes from preantral follicles to undergo embryogenesis and development to live young after growth, maturation, and fertilization *in vitro*. *Biol. Reprod.* **41**, 268–276 (1989).
9. Mehlmann, L. M. Oocyte-specific expression of Gpr3 is required for the maintenance of meiotic arrest in mouse oocytes. *Dev. Biol.* **288**, 397–404 (2005).
10. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* **101**, 6062–6067 (2004).
11. Zeng, F., Baldwin, D. A. & Schultz, R. M. Transcript profiling during preimplantation mouse development. *Dev. Biol.* **272**, 483–496 (2004).
12. Bleil, J. D. & Wassarman, P. M. Synthesis of zona pellucida proteins by denuded and follicle-enclosed mouse oocytes during culture *in vitro*. *Proc. Natl Acad. Sci. USA* **77**, 1029–1033 (1980).
13. Kasipillai, T. *et al.* Mutations in *EIF4ENIF1* are associated with primary ovarian insufficiency. *J. Clin. Endocrinol. Metab.* **98**, E1534–E1539 (2013).
14. Xu, X. L. *et al.* The microtubule-associated protein ASPM regulates spindle assembly and meiotic progression in mouse oocytes. *PLoS ONE* **7**, e49303 (2012).
15. Jiang, Z. Z. *et al.* Survivin is essential for fertile egg production and female fertility in mice. *Cell Death Dis.* **5**, e1154 (2014).
16. Hached, K. *et al.* Mps1 at kinetochores is essential for female mouse meiosis I. *Development* **138**, 2261–2271 (2011).
17. Adhikari, D. *et al.* Mastl is required for timely activation of APC/C in meiosis I and Cdk1 reactivation in meiosis II. *J. Cell Biol.* **206**, 843–853 (2014).
18. Salmon, E. D., Cimini, D., Cameron, L. A. & DeLuca, J. G. Merotelic kinetochores in mammalian tissue cells. *Phil. Trans. R. Soc. B* **360**, 553–568 (2005).
19. Cimini, D. Merotelic kinetochore orientation, aneuploidy, and cancer. *Biochim. Biophys. Acta* **1786**, 32–40 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the staff of the MRC Laboratory of Molecular Biology's Animal and Microscopy Facilities for technical assistance; J. Ellenberg, L. Jaffe and M. Freudzon for technical advice on culturing siRNA-injected follicles *in vitro*; J. Ellenberg and M. Madan Babu for discussions; G. Chalancon for help with data analyses; E. Voets for the *Mastl* construct; and members of the Schuh laboratory and S. Munro for comments on the manuscript. S.P. and M.P. received PhD fellowships from Boehringer-Ingelheim Fonds. T.T. is supported by an EMBO Long Term Fellowship (ALTF700-2014). The research leading to these results has received financial support from the European Research Council under grant agreement no. 337415 and the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 241548.

Author Contributions All authors analysed data. S.P. microinjected the majority of siRNA mixes, and identified, validated and characterized most genes; M.P. microinjected siRNA mixes, compared the developmental capacity and expression profile of oocytes grown *in vitro* and *in vivo*, and identified, validated and characterized several genes; V.K. microinjected siRNA mixes and wrote software in OriginPro to quantify phenotypes; T.T. validated and characterized *Dusp7*; B.S. did all bioinformatics analyses; V.K. and M.S. developed and established the strategy of the screen; M.S. wrote the manuscript; S.P., M.P., T.T. and B.S. commented on and edited the manuscript; M.P. and M.S. prepared the revised manuscript; M.S. supervised the study.

Author Information RNA sequencing data have been deposited in Gene Expression Omnibus under accession number GSE68150. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S. (mschuh@mrc-lmb.cam.ac.uk).

METHODS

Preparation, microinjection and culture of follicles. All mice were maintained in a specific pathogen-free environment according to UK Home Office regulations. Ovaries were dissected from two to five 10- to 12-day-old (C57BL × CBA) F₁ females. To obtain individual follicles, the ovaries were incubated in modified MEM- α (Gibco 12000-014) medium optimized for *in vitro* culture of follicles supplemented with 0.026 M NaHCO₃ (Sigma), 5678 U 100 ml⁻¹ penicillin G (Sigma) and 8265 U 100 ml⁻¹ streptomycin (Sigma), 1 × insulin/transferrin/selenium solution (ITS; Sigma; stock was 100 ×), 5% fetal bovine serum (FBS; Gibco 16000044) and 0.01 µg ml⁻¹ follicle stimulating hormone (FSH; National Hormone and Peptide Program, NDDK-oFSH-20) that was supplemented with 2 mg ml⁻¹ collagenase (Roche) for about 30–40 min total. During incubation with collagenase, the ovaries were pipetted up and down every 10 min to facilitate dissociation and then washed through several droplets of follicle culture medium without collagenase. The follicles were then randomly allocated into control siRNA and RNAi mix injection groups. Intact follicles were then loaded into a microinjection chamber prepared with two double stick tapes as spacer and microinjected as previously described⁷ in culture medium supplemented with HEPES (Sigma). Upon microinjection, follicles were cultured at 37 °C in 5% CO₂ on membrane inserts in 6- or 12-well culture dishes filled with follicle culture medium (see above). For most experiments, collagen-coated inserts from Corning were used (Transwell COL), but also Transwell-Clear inserts that were coated with 10 µg cm⁻² collagen solution type I from rat tail (Sigma), BD Matrigel Basement Membrane (BD Biosciences; thin coating method) as well as BD BioCoat filters were successfully used (Extended Data Fig. 1c). Medium surrounding the filter was replaced with fresh medium every 3–4 days. Oocytes were isolated from follicles after 10–11 days of *in vitro* culture. To this end, the oocytes were stripped with a small glass pipette and released into modified M2 medium that contained 10% FBS instead of BSA as well as 100 ng ml⁻¹ FSH and dbcAMP. Oocytes were subsequently microinjected with mRNAs encoding eGFP- α -tubulin (spindle) and H2B-mRFP (chromosomes). Upon microinjection, oocytes were cultured for up to 3.5 h at 37 °C until fluorescent proteins were expressed. Oocytes were then released from prophase arrest by transferring them into medium without dbcAMP. *In vivo* grown control oocytes from 5-week-old (C57BL × CBA) F₁ females were obtained by puncturing isolated ovaries with hypodermic needles and then microinjected with mRNAs as described above.

***In vitro* fertilization.** Oocytes grown for 10–11 days within follicles *in vitro* or obtained from adult (C57BL × CBA) F₁ females (7–12 weeks old) were denuded and matured in follicle-culture medium (see above). Meiosis II oocytes were placed in 50 µl of EmbryoMax HTF medium (Millipore) and fertilized with 10 µl of sperm suspension from 5- to 13-week-old (C57BL × CBA) F₁ males. The sperm suspension was prepared by dissecting two cauda epididymus from one male in 2 ml of HTF medium. After 4–6 h, zygotes were transferred to KSOM+AA (Millipore) and cultured for 5 days at 37 °C.

Expression constructs and mRNA synthesis. To generate constructs for *in vitro* mRNA synthesis, the previously published protein coding sequences of *Mastl*²⁰ and α -tubulin²¹ were fused with eGFP and inserted into pGEMHE for *in vitro* transcription. The *Mus musculus Dusp7* ORF (derived from NM_153459) was amplified by PCR from mouse oocyte cDNA. The resulting product of around 1,300 base pairs contained a 5'-XhoI and a 3'-EcoRI restriction site, which were used to insert it into pGEMHE-eGFP carboxy terminally of the eGFP tag. These constructs, as well as pGEMHE-H2B-mRFP1²², pGEMHE-eGFP-MAP4²² and pGEMHE-eGFP-LaminB1²³, were linearized with AscI. Capped mRNA was synthesized using T7 polymerase (mMessage mMachine kit, following the manufacturer's instructions, Ambion) and dissolved in 11 µl water. mRNA concentrations were determined on ethidium bromide agarose gels by comparison with an RNA standard (Ambion).

Confocal microscopy. Images were acquired with a Zeiss LSM710 confocal microscope equipped with a Zeiss environmental incubator box or a Zeiss LSM780 confocal microscope equipped with a Tokai Hit Stage Top Incubator, with a C-Apochromat 40×/1.2 W water immersion objective lens for live oocytes, and a C-Apochromat 63×/1.2 W water immersion objective for fixed oocytes as previously described²². In some images, shot noise was reduced with a Gaussian filter. The z-projections were generated in Zen (Zeiss).

Measurement of chromosome volumes. Oocytes were fixed for 30 min at 37 °C in 100 mM HEPES (pH 7) (titrated with KOH), 50 mM EGTA (pH 7) (titrated with KOH), 10 mM MgSO₄, 2% formaldehyde (MeOH free) and 0.2% Triton X-100. DNA was stained with 0.05 µg ml⁻¹ Hoechst 33342 (Molecular Probes). All stainings were performed in PBS, 0.1% Triton X-100, 3% BSA. Chromosome volumes were determined in three-dimensional volume reconstructions using the surface function in Imaris (Bitplane).

All siRNAs were purchased from QIAGEN. siRNAs were diluted in 96-well plates to a concentration of 6.6 µM and stored at -80 °C. To pre-select genes, we took advantage of two microarray data sets^{10,11}, which compare the expression profile of oocytes with the profiles of other cell types and preimplantation mouse embryos, respectively. Only genes that were highly significantly upregulated in oocytes in both data sets were selected for the screen, independently of whether they had previously been implicated in mitosis or meiosis to avoid any bias. We targeted each gene with a low complexity siRNA pool (three siRNAs per gene) (Supplementary Table 4), which on average leads to fewer off-target effects and a higher penetrance of phenotypes than individual siRNAs^{24,25}. For the primary RNAi screen, siRNAs targeting different genes were mixed and microinjected to a final concentration of 5 nM each in the oocyte. For the functional characterization of individual genes, siRNA concentrations of up to 0.2 µM final in the oocyte were used. Protein ablation should always be assessed by secondary assays, because proteins that are generated in the very early stages of meiosis may still not be efficiently depleted if they do not turn over, even if the targeted transcript is reduced.

Quantitative real-time PCR. mRNA was extracted using an RNeasy Mini Kit (Qiagen) and cDNA was generated using a High Capacity RNA-to-cDNA Kit (Applied Biosystems). Real-time PCR was performed with the 7900 HT Real-Time Fast PCR System (Applied Biosystems) using SYBR Green. *GAPDH* mRNA was used for normalization.

RNA sequencing. Total RNA was isolated using NucleoSpin RNA XS (Macherey-Nagel) from oocytes grown *in vitro* after 10 days of follicle culture or from oocytes obtained directly from adult (C57BL × CBA) F₁ females (7–11 weeks old). A total of 50 oocytes with an intact nucleus per sample were used and three samples per group were collected. RNA was extracted using NucleoSpin RNA XS (Macherey-Nagel). A cDNA library was prepared using SMARTer UltraLow Input RNA for Sequencing (Clontech Laboratories) and the samples were processed by BGI Tech Solutions. The cDNA product was synthesized and amplified using a SMARTer PCR cDNA Synthesis Kit (Clontech Laboratories) from the total RNA (10 ng) of each sample. The cDNA was fragmented by Covaris E210 and the median insert length was about 200 base pairs. The paired-end cDNA library was prepared in accordance with Illumina's protocols with an insert size of 200 base pairs and sequenced for 100 base pairs by HiSeq2000 (Illumina).

Expression analysis. NOISeq. RNA-Seq based measurements of transcript abundances at the level of genes were represented by fragments per kilobase of transcript per million fragments mapped (FPKM). FPKM is conceptually similar to the reads per kilobase per million reads sequenced (RPKM) measure, but it is easily adaptable for sequencing data from one to higher numbers of reads from single source molecules. To identify significantly differentially expressed genes between oocytes grown *in vitro* and *in vivo*, we used a non-parametric method encoded in NOISeq. For this, we first filtered for low count or abundance using the 'CPM' low count filter of NOISeq. This yielded a reduction from the original 16,343 genes to 11,470 genes. Significant differential expression between oocytes grown *in vivo* and *in vitro* was determined using NOISeq with the following parameters: (1) 'tmm', trimmed mean of log₂ FPKM, normalization; (2) biological replicates data; and (3) probability of differential expression *q* being set to 0.8 or above and log₂ values being greater than or equal to -1 or 1 for upregulated and downregulated genes respectively in the *in vitro* group. This yielded 146 upregulated and 67 downregulated genes in oocytes grown *in vitro*. The vast majority of genes (11,110) were unchanged between the two conditions.

DESeq2. RNA-seq counts were considered with two 'conditions', namely *in vivo* and *in vitro* with three replicates. The standard protocol for DESeq2 differential expression analyses was followed with default settings. We deemed genes to be upregulated or downregulated if log₂ values were greater than or equal to -1 or 1 for upregulated and downregulated genes respectively in the *in vitro* group, with a false discovery rate (FDR or padj of DESeq2) less than 0.01 or 1%. We considered a particularly low value of false discovery rate because of overall low expression levels for transcripts. Hence, we used a more stringent value for the false discovery rate. This yielded 282 upregulated and 163 downregulated genes *in vitro*.

Statistics. Mean, s.d. and statistical significance based on Student's *t*-test or Fisher's exact test (two-tailed) were calculated in Microsoft Excel, assuming normal distribution and similar variance. No statistical methods were used to pre-determine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. All error bars show s.d. All box plots show median (line), mean (small square), 5th, 95th (whiskers) and 25th and 75th percentile (boxes). The *z*-scores were calculated as the deviation of the mean of a single siRNA mix to the mean of all controls of the RNAi screen, normalized to the s.d. of all controls. siRNA mixes were sorted according to their *z*-score. The dashed line in Fig. 2 and Extended Data Fig. 4 delineates mixes with a *z*-score higher than two s.d. above the average value of all controls.

Data analysis. Phenotypes were evaluated manually by browsing the data in Zen (Zeiss). Defects and measurements (time points and spindle parameters) were then recorded on a homemade user interface in OriginPro 8.0 and processed in Microsoft Excel. Averages, s.d. and statistical significance were calculated in Excel. The z-scores were calculated as the deviation of the mean of a single mix to the overall mean of all controls, normalized to the s.d. of all controls. Oocytes that died during imaging were not analysed and do not contribute to the data set.

For Fig. 4c, we analysed data from all 2,241 oocytes, because lagging chromosomes are not very common in control oocytes, but are likely to be triggered by various defects such as those induced by RNAi in the screen.

For the Jaccard index heatmap in Extended Data Fig. 9, RNAi screen phenotypes from both mix and control experiments were collected; wherever there were numerical values, they were converted appropriately into 'yes' and 'no' values based on the mean and s.d. of the distribution of numerical values. Further, 'yes' values were categorized into '+' and '-' groups based on whether a numerical entry was greater than mean + s.d. or smaller than mean - s.d. This information was converted to a network representation such that there were two types of node in the network oocytes and phenotypes (Extended Data Figure 9c). An edge was made between oocyte and phenotype if a given oocyte scored 'yes' for a given phenotype. This yielded a network that we termed phenotype-oocyte network, which included 5,203 edges (or associations) between 53 phenotypes and 1,504 oocytes. The distribution of the number of oocytes against the number of distinct phenotypes scored in them suggested that over 75% of oocytes, namely 1,195, have two or more phenotypes scored, suggesting that there were widespread multiple phenotypes scored for in the vast majority of oocytes, as expected. Hence, we sought to estimate the extent of co-occurring phenotypes across oocytes as a first step towards phenotype correlations. We calculated the Jaccard index between all possible pairs of phenotypes in the phenotype-oocyte network. The Jaccard index between phenotype i and phenotype j was defined as

$$\frac{\text{oocytes exhibiting phenotype } i \cap \text{oocytes exhibiting phenotype } j}{\text{oocytes exhibiting phenotype } i \cup \text{oocytes exhibiting phenotype } j}$$

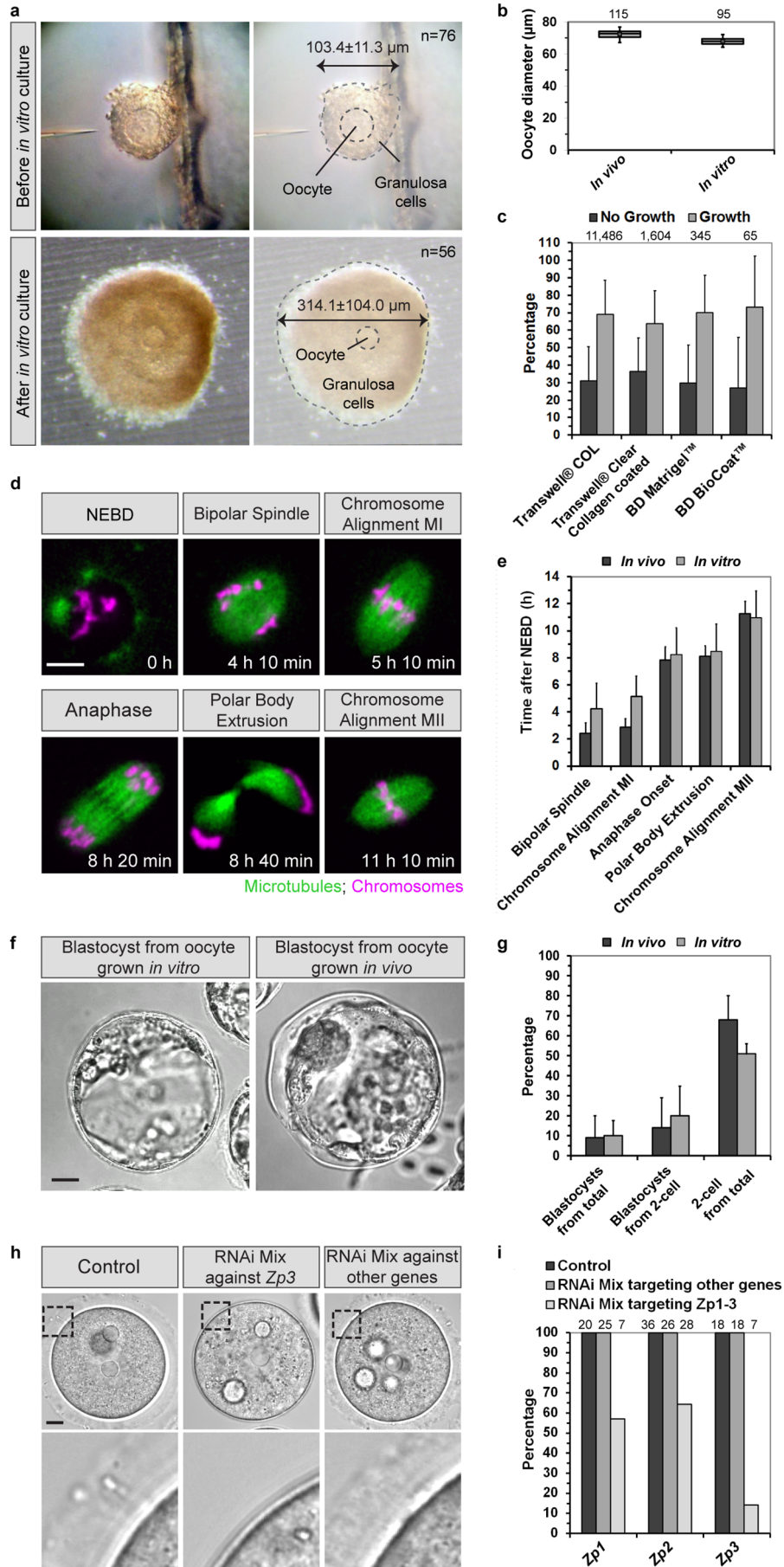
where \cap denotes 'intersection' between sets of oocytes with phenotypes i, j and \cup denotes 'union' between sets of oocytes with phenotypes i, j .

The above formula for the Jaccard index captures the fraction of co-occurrence of phenotypes i and j in oocytes over the total observed number of instances of phenotypes i or j . The numerator denotes the number of oocytes in which phenotypes i and j were observed, while the denominator indicates the total number of oocytes in which either phenotypes i or j have been observed. The values of a Jaccard index range between 0 and 1. Zero signifies poor co-occurrence while '1' signifies

high co-occurrence. We calculated the Jaccard index for all possible pairs of phenotypes in the phenotype-oocyte network. Out of a possible 1,378 ($53C_2$), we could obtain 844 pairs that displayed a Jaccard index greater than zero. We then clustered the profile of the Jaccard index between phenotypes represented as a matrix or table. For this purpose, we used pheatmap (<http://cran.r-project.org/web/packages/pheatmap/index.html>) with 'mean' clustering and 'Pearson correlation' options. In this way, we obtained three major clusters of phenotype correlations.

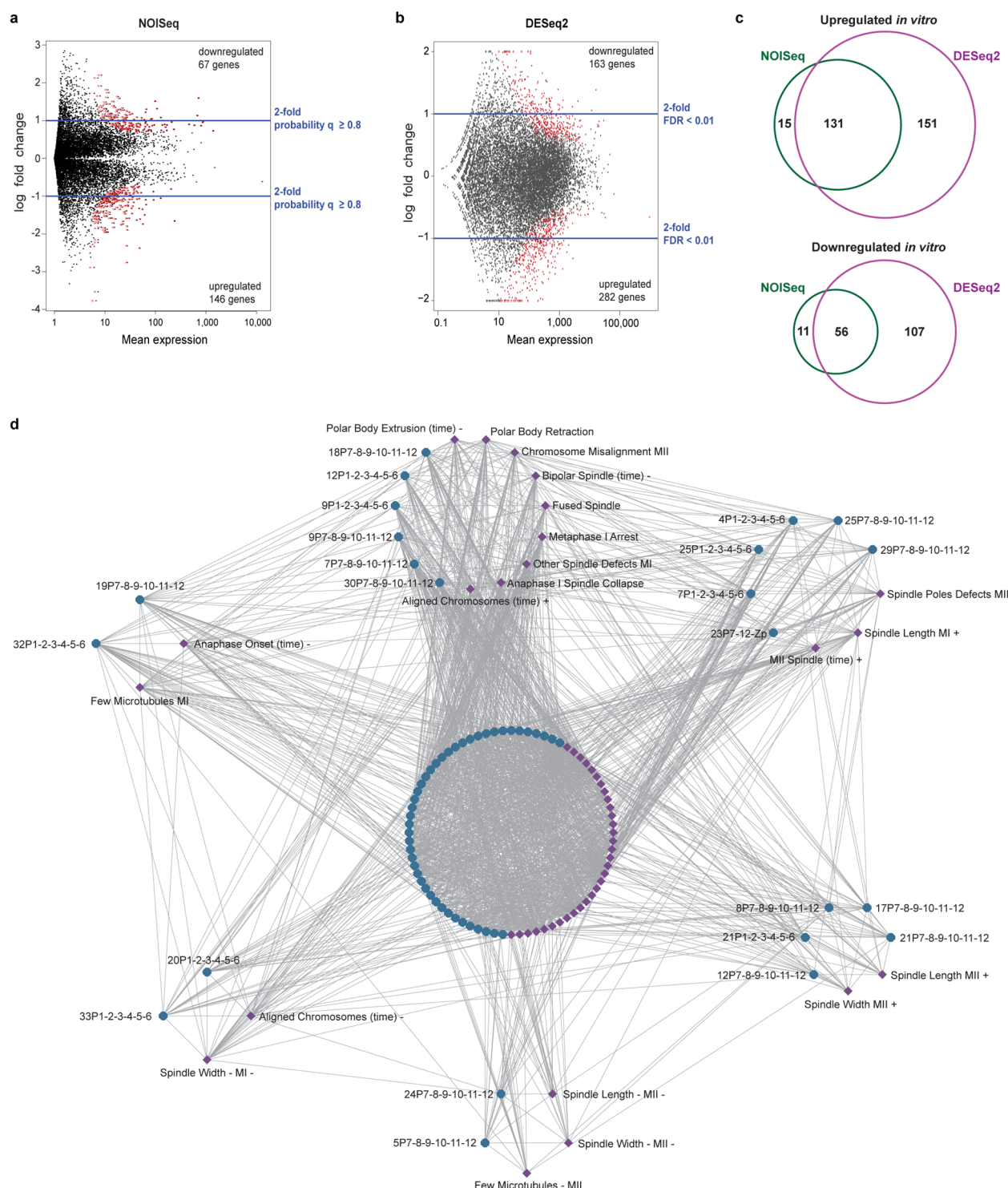
Measurement of oocyte diameter, spindle length and spindle width. Oocyte diameter, spindle length and width in metaphase I and metaphase II were measured using the Measurement function in Zen (Zeiss). To measure the oocyte diameter accurately, measurements were always taken in the centre of the oocyte as determined by the maximum radius of the oocyte. Spindle length and width were only measured in oocytes in which the spindle was parallel to the confocal imaging plane.

20. Voets, E. & Wolthuis, R. M. MASTL is the human orthologue of Greatwall kinase that facilitates mitotic entry, anaphase and cytokinesis. *Cell Cycle* **9**, 3591–3601 (2010).
21. Mora-Bermudez, F., Gerlich, D. & Ellenberg, J. Maximal chromosome compaction occurs by axial shortening in anaphase and depends on Aurora kinase. *Nature Cell Biol.* **9**, 822–831 (2007).
22. Schuh, M. & Ellenberg, J. Self-organization of MTOCs replaces centrosome function during acentrosomal spindle assembly in live mouse oocytes. *Cell* **130**, 484–498 (2007).
23. Lenart, P. *et al.* Nuclear envelope breakdown in starfish oocytes proceeds by partial NPC disassembly followed by a rapidly spreading fenestration of nuclear membranes. *J. Cell Biol.* **160**, 1055–1068 (2003).
24. Echeverri, C. J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nature Rev. Genet.* **7**, 373–384 (2006).
25. Parsons, B. D., Schindler, A., Evans, D. H. & Foley, E. A direct phenotypic comparison of siRNA pools and multiple individual duplexes in a functional assay. *PLoS ONE* **4**, e8471 (2009).
26. Griffin, J., Emery, B. R., Huang, I., Peterson, C. M. & Carrell, D. T. Comparative analysis of follicle morphology and oocyte diameter in four mammalian species (mouse, hamster, pig, and human). *J. Exp. Clin. Assist. Reprod.* **3**, 2 (2006).
27. Ge, L. *et al.* Factors affecting the in vitro action of cumulus cells on the maturing mouse oocytes. *Mol. Reprod. Dev.* **75**, 136–142 (2008).
28. Zhou, P. *et al.* Mouse cumulus-denuded oocytes restore developmental capacity completely when matured with optimal supplementation of cysteamine, cystine, and cumulus cells. *Biol. Reprod.* **82**, 759–768 (2010).
29. Alvarez-Fernandez, M. *et al.* Greatwall is essential to prevent mitotic collapse after nuclear envelope breakdown in mammals. *Proc. Natl Acad. Sci. USA* **110**, 17374–17379 (2013).
30. Wang, P. *et al.* Cell cycle regulation of Greatwall kinase nuclear localization facilitates mitotic progression. *J. Cell Biol.* **202**, 277–293 (2013).



Extended Data Figure 1 | Efficiency of follicle growth and comparison of oocytes grown *in vitro* and *in vivo*. **a**, Follicles before (top panel) and after (bottom panel) *in vitro* culture. The perimeters of oocyte and granulosa cells are highlighted on the right. The follicle diameter increases from $103.4 \pm 11.3 \mu\text{m}$ to $314.1 \pm 104.0 \mu\text{m}$ during *in vitro* culture. This lies between the diameter of *in vivo* grown early antral ($\sim 248 \mu\text{m}$) and Graafian ($\sim 424 \mu\text{m}$) mouse follicles²⁶. The diameter of *n* follicles was measured before and after *in vitro* culture and is displayed as mean \pm s.d. Measurements from three or two independent experiments for before and after culture, respectively. **b**, Diameter of oocytes grown *in vivo* or *in vitro*. Data from two and seven experiments, respectively. The box plot shows median (line), mean (small square), 5th, 95th (whiskers) and 25th and 75th percentile (boxes). **c**, Efficiency of follicle growth on different culture substrates. The numbers of independent experiments are 343, 56, 11 and 3 from left to right. The total number of follicles is specified above the bars. Error bars, s.d. **d**, Live oocyte expressing eGFP-MAP4 (green, microtubules) and H2B-mRFP (magenta, chromosomes). The characteristic time points of oocyte maturation that were determined for each oocyte in the screen (2,241 oocytes in total from 70 experiments) are listed above the representative images. Quantification of timing in **e**. Scale bar, $10 \mu\text{m}$. **e**, The timing of bipolar spindle assembly, chromosome alignment during meiosis I, anaphase, polar body extrusion and chromosome alignment during meiosis II were quantified in oocytes obtained from 5-week-old (C57BL \times CBA) F₁ females or in oocytes from the same

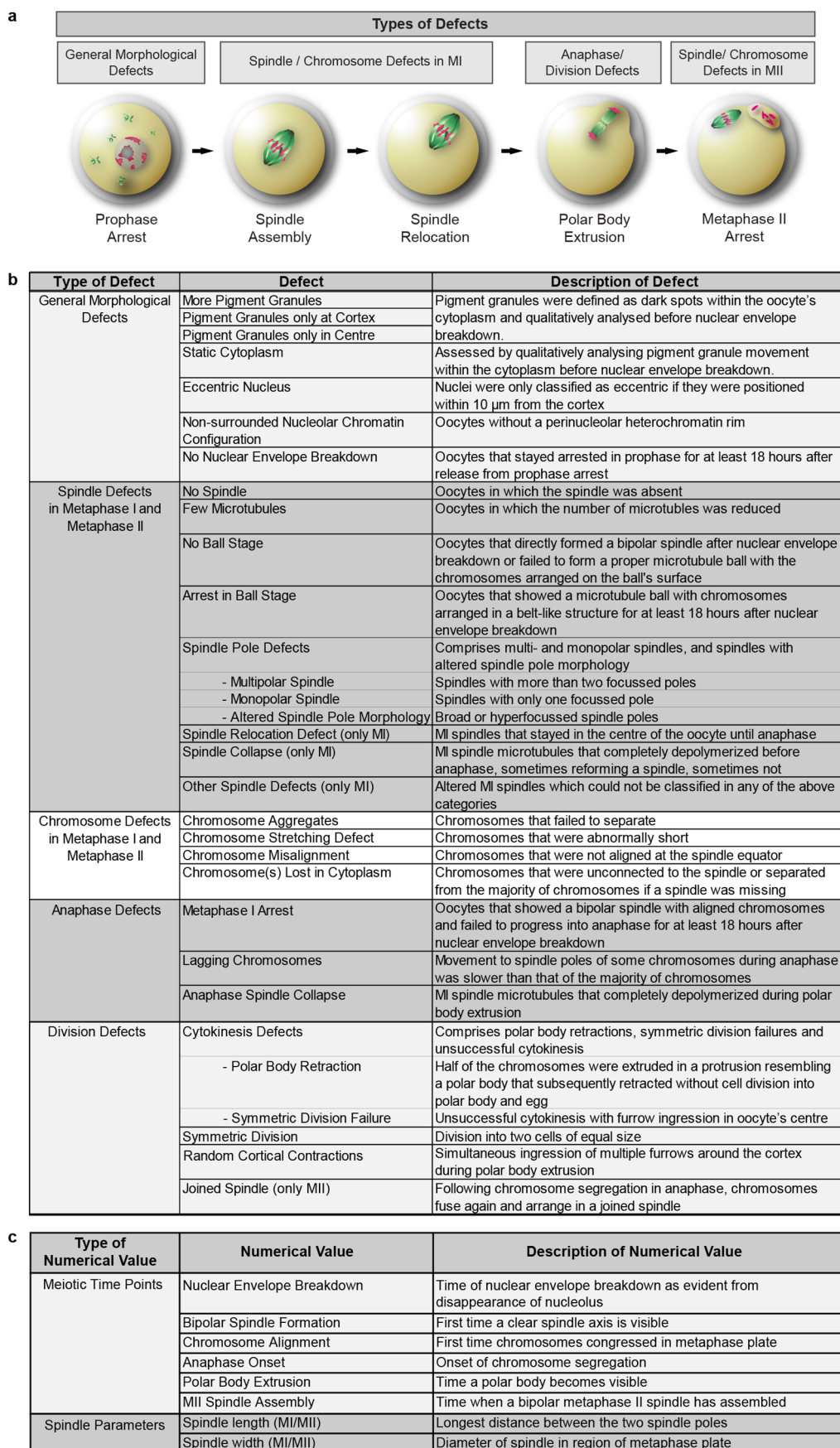
strain grown *in vitro* within follicles. Data from four independent experiments. Error bars, s.d. **f**, Transmitted light images of blastocysts derived from fertilized (C57BL \times CBA) F₁ oocytes grown *in vitro* within follicles (left) or *in vivo* (right). Scale bar, $20 \mu\text{m}$. Quantifications in **g**. **g**, (C57BL \times CBA) F₁ oocytes grown *in vitro* within follicles or *in vivo* were denuded, matured *in vitro* and fertilized. The percentages of all oocytes (fertilized and unfertilized) that developed into two-cell embryos (two-cell from total) and two-cell embryos that developed into blastocysts (blastocyst from total) were quantified. Developmental rates are consistent with previous studies, in which *in vitro* matured denuded oocytes were fertilized^{27,28}; 179 oocytes grown *in vivo* and 180 oocytes grown *in vitro* were analysed in total. Data from three independent experiments for each group. Error bars, s.d. **h**, Transmitted light images of control oocytes and oocytes microinjected with an siRNA mix targeting *Zp3* together with 11 other genes (RNAi Mix against *Zp3*) or an siRNA mix microinjected at the same time that targeted 12 other genes (RNAi mix against other genes). Highlighted region is magnified below. Scale bar, $10 \mu\text{m}$. Quantification of phenotype in **i**. **i**, The presence of the zona pellucida was scored in oocytes microinjected with control siRNA (control), an siRNA mix targeting one of the three *Zp* genes (*Zp1*, *Zp2* or *Zp3*) together with 11 other genes and an siRNA mix microinjected at the same time that targeted 12 different genes (RNAi mix against other genes). The number of analysed oocytes is given next to bars.



Extended Data Figure 2 | Transcriptome analysis of oocytes grown *in vivo* and *in vitro*. **a–c**, Transcriptome analysis of oocytes grown *in vitro* and *in vivo*.

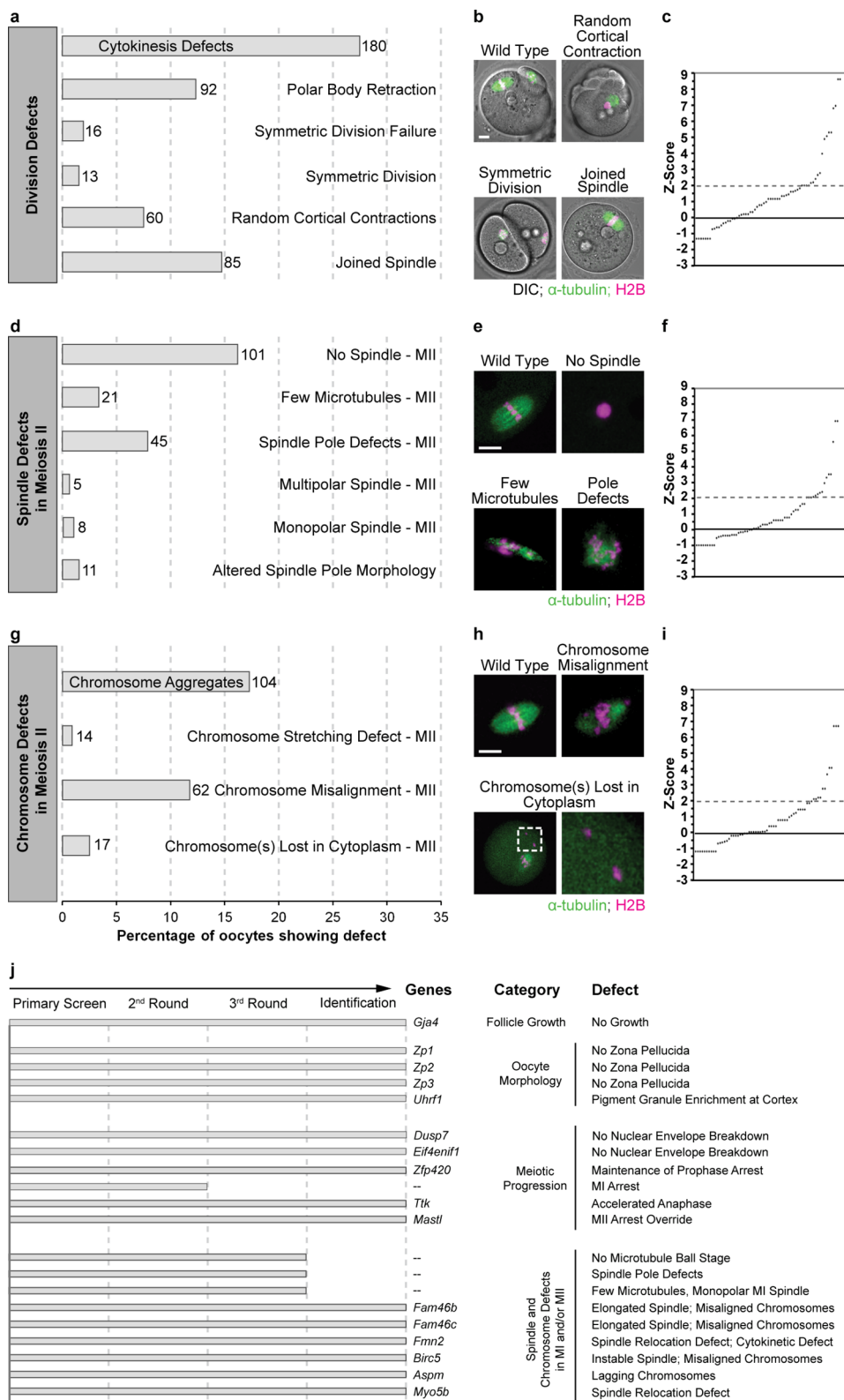
a, Differentially expressed genes in oocytes grown *in vitro* based on evaluation using NOISeq algorithm. Transcript abundances are reported in transcript FPKM. Only about 2% (213 out of 11,470) of genes were differentially expressed. **b**, Differentially expressed genes in oocytes cultured *in vitro* based on evaluation using DESeq2 algorithm. Only about 4% (445 genes out of 10,597) of genes were differentially expressed after applying filters in both **b** and **c**. The blue lines indicate genes with at least twofold change in expression. Red colour indicates differentially expressed genes with the denoted probability. For details, see Methods. **c**, The overlap between NOISeq and

DESeq2 results, presented as Venn diagrams. There is at least over 80% overlap in genes in either upregulated or downregulated groups for both NOISeq and DESeq2. **d**, Qualitative network of phenotypes in oocytes microinjected with siRNA mixes. Blue nodes represent siRNA mixes, purple nodes represent phenotypes. Grey lines between mixes and phenotypes denote if at least one oocyte microinjected with a given mix displayed the phenotype. The clusters indicate a close relationship between a set of phenotypes and mixes. The clusters were obtained using ClusterViz (<https://code.google.com/p/clusterviz-cytoscape/>) of Cytoscape, which encodes the MCODE method to identify clusters of closed related nodes based on the topology of the network. The network contains six clusters identified by ClusterViz.



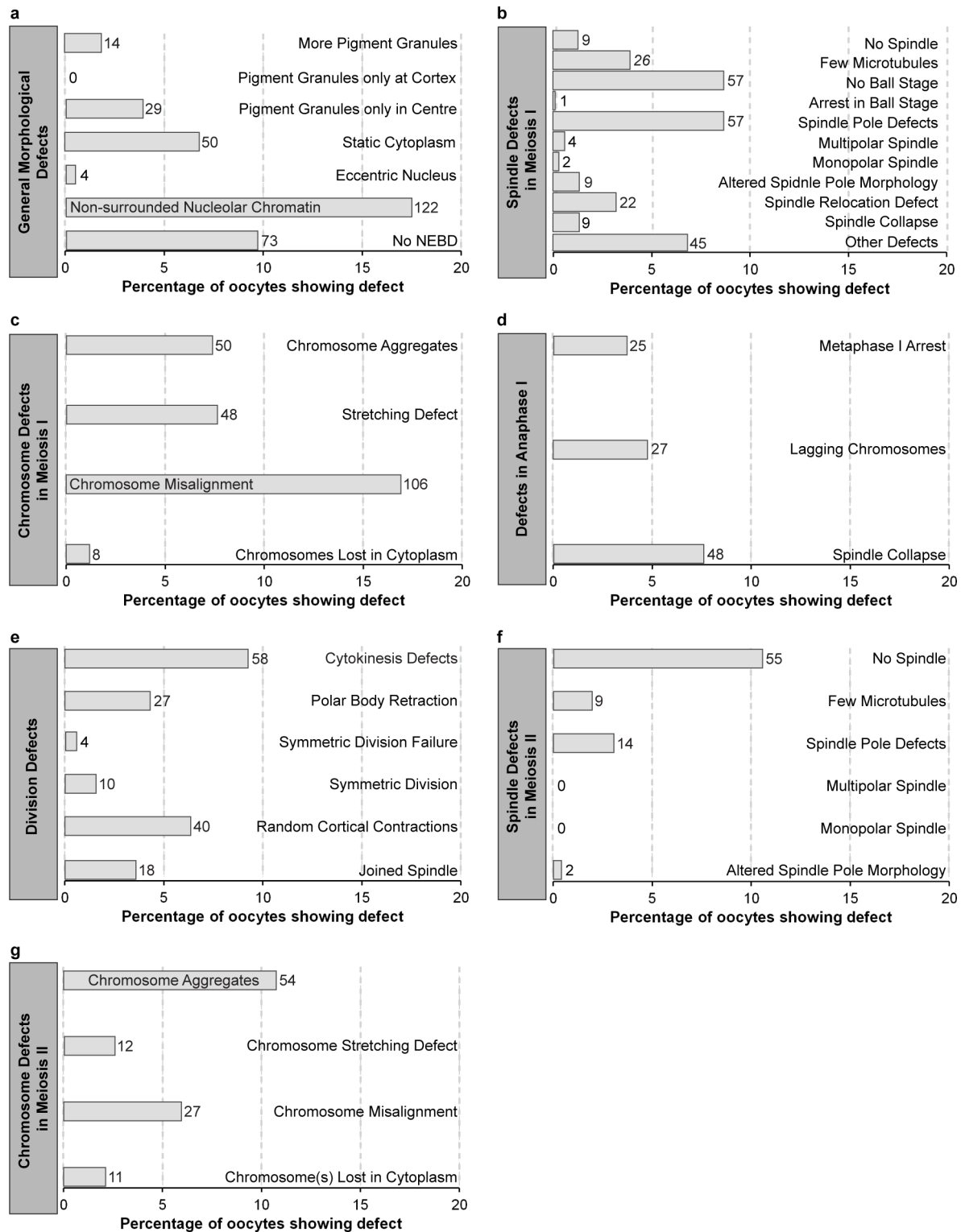
Extended Data Figure 3 | Description of defects scored in screen. **a**, Scheme illustrating the main categories of defects that were quantified in the screen. **b**, Table listing the main categories of defects and their subcategories as well as a

description of each defect. **c**, Table listing the numerical values that were measured in the screen and a description of each numerical value.



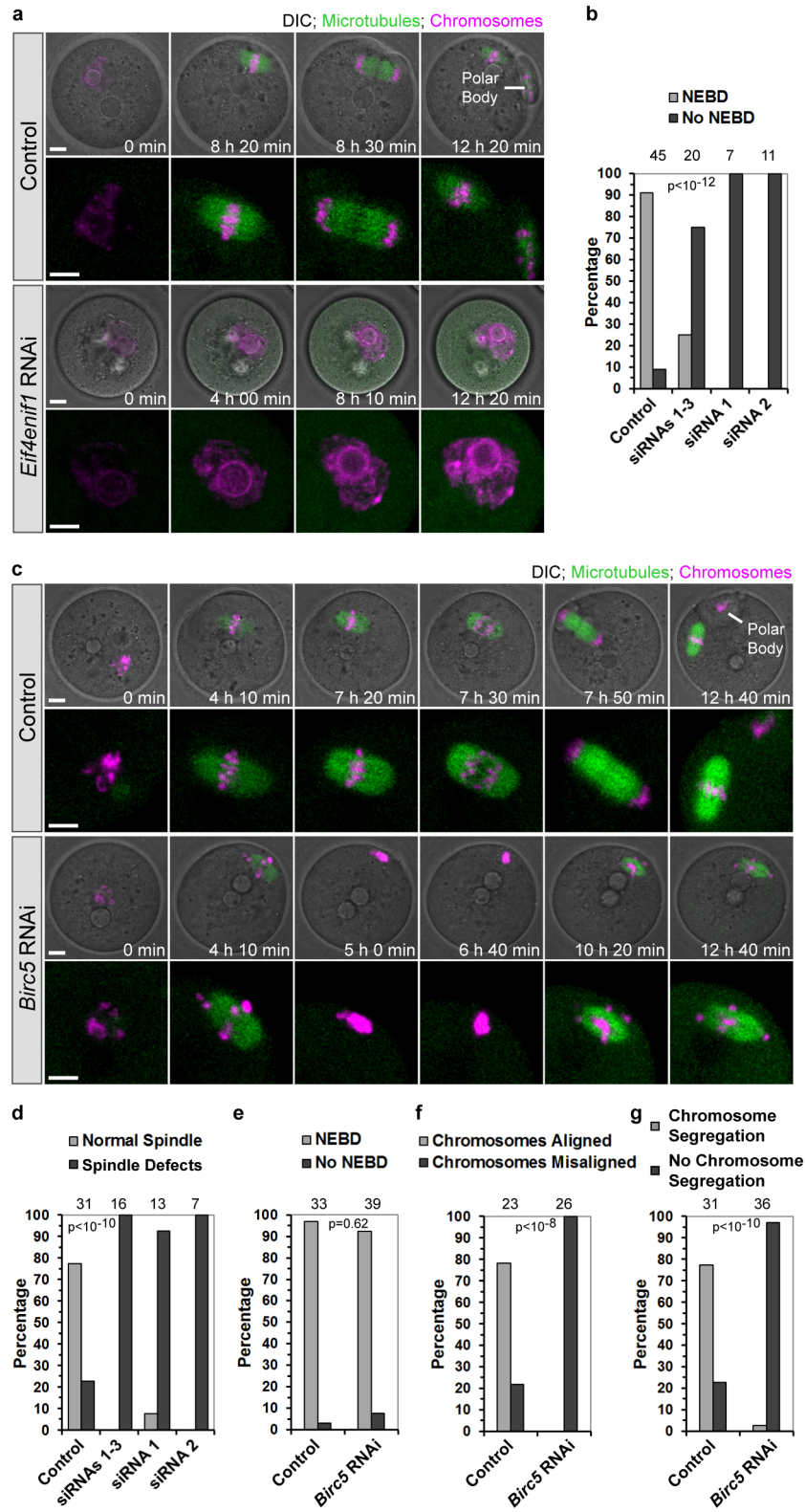
Extended Data Figure 4 | Defects during meiosis II in siRNA-treated oocytes. **a, d, g,** The frequency of cytokinetic defects (**a**), spindle defects in metaphase II (**d**) and chromosome defects in metaphase II (**g**) were scored in siRNA-treated oocytes. The absolute number of oocytes with each defect is given next to bars. Data from 70 independent experiments. Corresponding control data are shown in Extended Data Fig. 5. **b, e, h,** Examples of defects in live oocytes. Chromosomes (magenta) were labelled with H2B-mRFP, microtubules (green) with eGFP- α -tubulin. Quantifications in **a, d, g.** Scale bars, 10 μ m. **c, f, i,** The z-scores were calculated as the deviation of the mean of a single siRNA

mix to the mean of all controls of the RNAi screen, normalized to the s.d. of all controls. siRNA mixes were sorted according to their *z*-score. The dashed line delineates mixes with a *z*-score higher than two s.d. above the average value of all controls. **j**, List of genes that were tracked down to the individual gene level in the RNAi screen. Note that defects caused by depletion of some proteins such as *Zfp420* or *Uhrf1* may reflect the function of more proximal genes under the control of these proteins. We were able to allocate 16 out of 20 tested defects to individual genes. Defects that could not be tracked down to individual gene level are shown as grey bars ending after the second or third round.



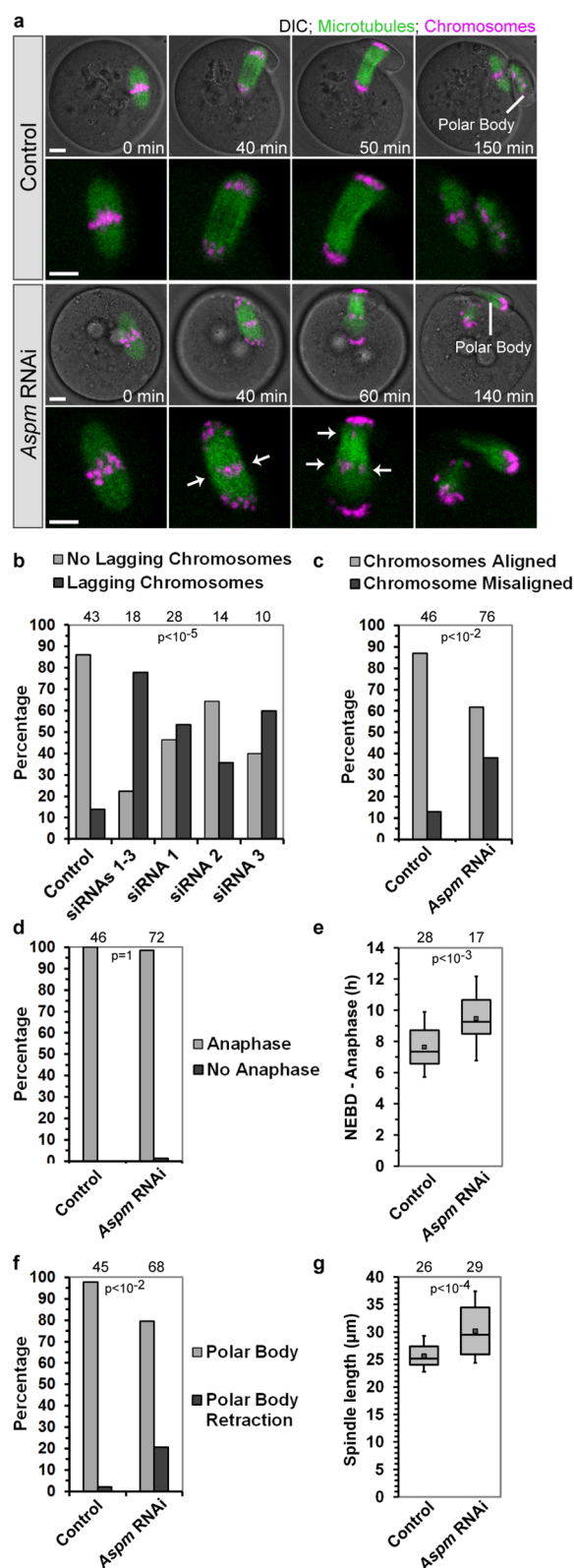
Extended Data Figure 5 | Frequency of meiosis I and meiosis II defects in oocytes treated with control siRNAs. a–g. The frequency of scored general morphological defects (a), spindle defects in meiosis I (b), chromosome defects in meiosis I (c), defects in anaphase I (d), defects during cytokinesis (e),

spindle defects in meiosis II (f) and chromosome defects in meiosis II (g) were scored in oocytes microinjected with control siRNAs. The absolute number of oocytes with each defect is given next to bars.



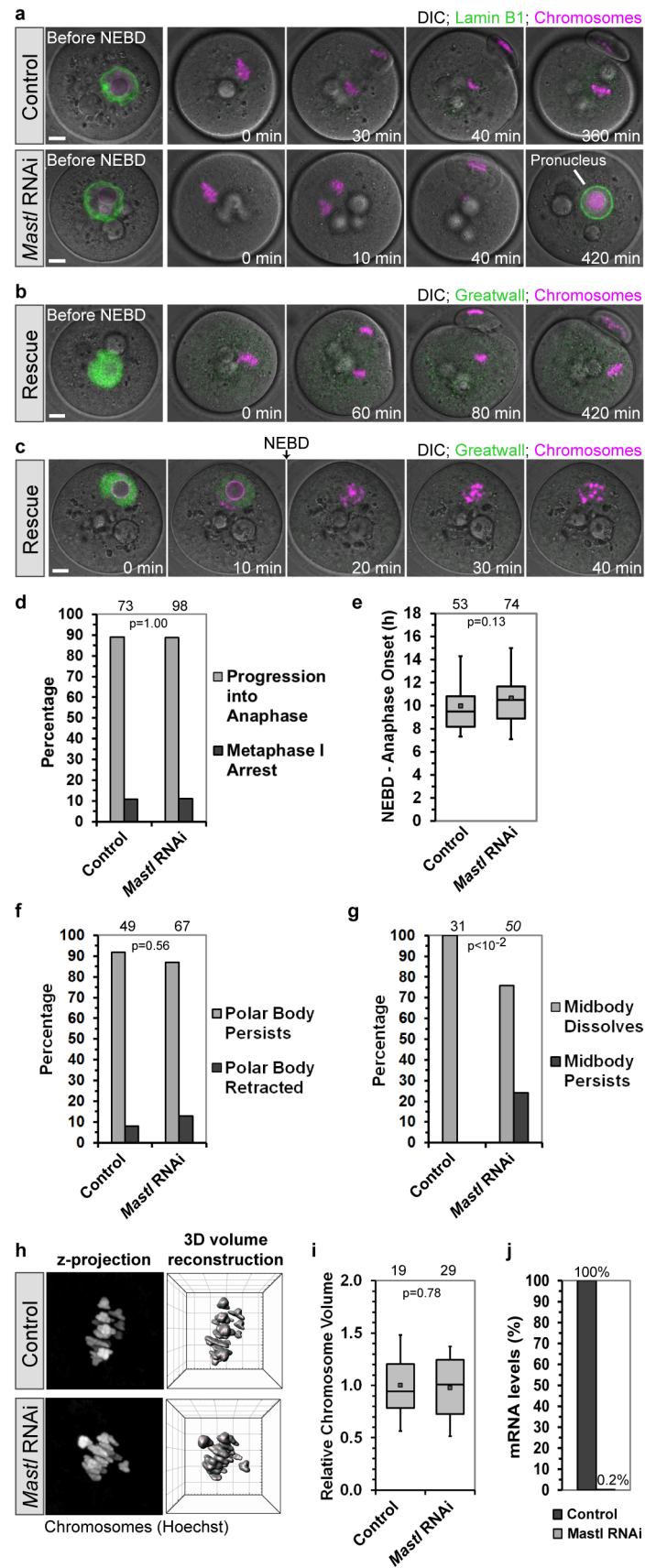
Extended Data Figure 6 | *Eif4enif1* is required for release from prophase arrest and *Birc5* for spindle integrity. **a**, Live oocytes microinjected with control siRNA (control) or siRNAs targeting *Eif4enif1* (*Eif4enif1* RNAi) expressing eGFP- α -tubulin (green, microtubules) and H2B-mRFP (magenta, chromosomes) merged with differential interference contrast (DIC) image. Region of spindle and chromosomes is magnified without DIC below. Quantification of phenotype in **b**. Scale bar, 10 μ m. **b**, Live oocytes microinjected with control siRNA or *Eif4enif1* siRNAs were monitored by long-term time-lapse microscopy as shown in **a** and the efficiency of NEBD was scored. The number of analysed oocytes is specified next to bars. The *P* value was calculated with Fisher's exact test. Data from a total of three experiments. **c**, Live oocytes microinjected with control siRNA (control) or siRNAs targeting *Birc5* (*Birc5* RNAi) expressing eGFP- α -tubulin (green, microtubules) and H2B-mRFP (magenta, chromosomes) merged with DIC. Region of spindle and

chromosomes is magnified without DIC below. Quantification of phenotypes in **d–g**. Scale bar, 10 μ m. **d**, Live oocytes microinjected with control siRNA (control), a mix of three different *Birc5* siRNAs (siRNA 1–3) or two *Birc5* siRNAs individually (siRNA 1, 2) were scored for temporary or permanent disintegration of the meiotic spindle. The number of analysed oocytes is specified next to bars. The *P* value was calculated with Fisher's exact test comparing control and all *Birc5* siRNA microinjected oocytes from five experiments. **e–g**, Live oocytes microinjected with control siRNA or *Birc5* siRNAs were monitored by long-term time-lapse microscopy as shown in **c** and the efficiency of NEBD (**e**), the presence or absence of misaligned chromosomes (**f**) as well as the efficiency of chromosome segregation (**g**) were scored. The number of analysed oocytes is specified next to bars. *P* values were calculated with Fisher's exact test. Data (**d–g**) from five independent experiments.



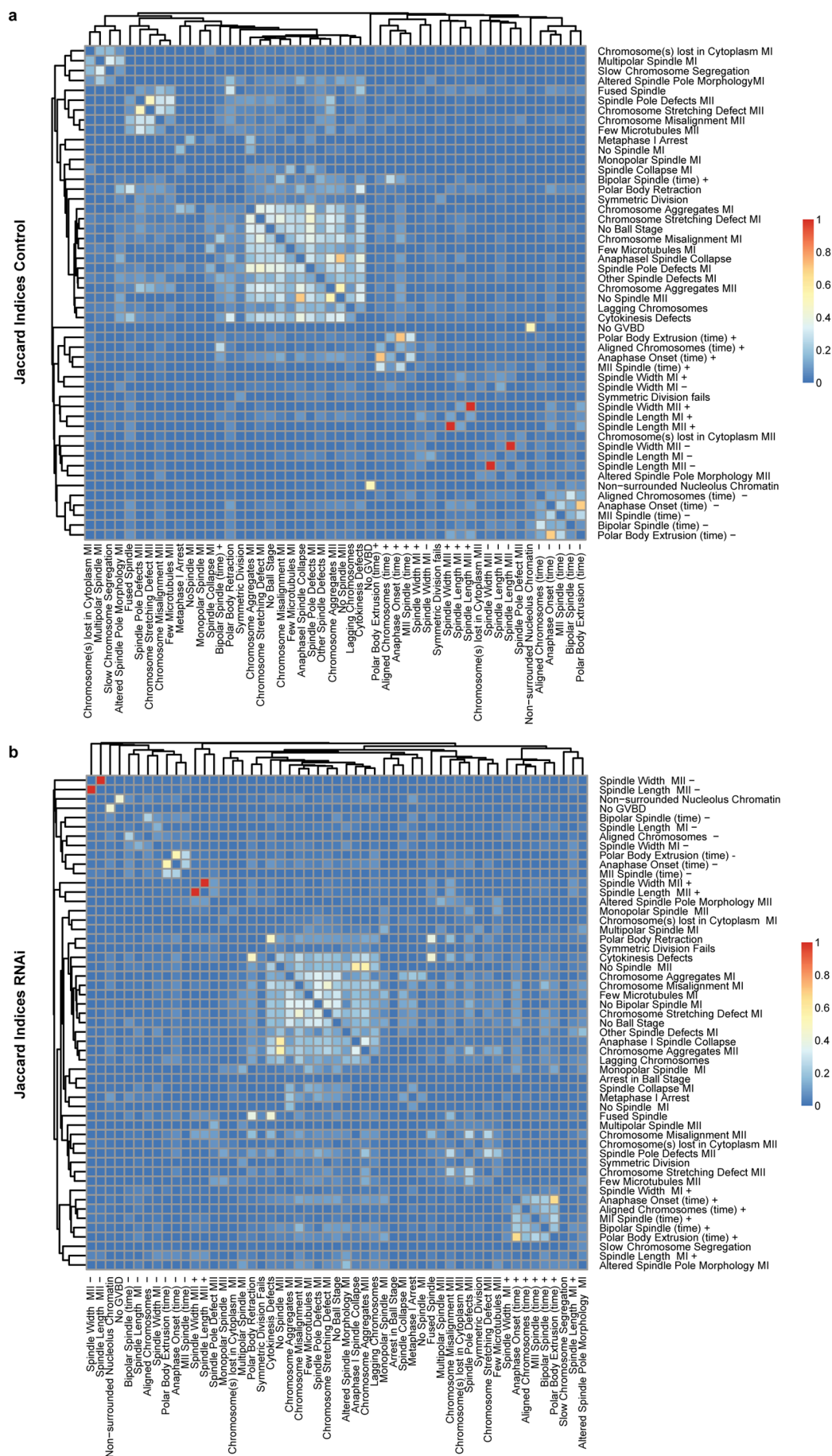
Extended Data Figure 7 | *Aspm* function in mouse oocytes. **a**, Oocytes microinjected with siRNAs targeting *Aspm* or injected with control siRNA. Microtubules in green, chromosomes in magenta. Arrows highlight lagging chromosomes. Quantification of phenotypes in **b–g**. Scale bar, 10 μm. **b, c**, Lagging (**b**) or misaligned chromosomes (**c**) in oocytes microinjected with different *Aspm* siRNAs. **d–g**, Live oocytes microinjected with control siRNA (control) or *Aspm* siRNAs (*Aspm* RNAi) were monitored by long-term time-lapse microscopy as shown in **a** and scored for progression through

anaphase (**d**), time of anaphase onset (**e**), polar body extrusion (**f**) and spindle length (**g**). The number of analysed oocytes is specified next to bars. The *P* value was calculated with Fisher's exact test (**b, c, d, f**) or Student's *t*-test (**e, g**) comparing control and all *Aspm* siRNA microinjected oocytes. The box plots in **e** and **g** show median (line), mean (small square), 5th, 95th (whiskers) and 25th and 75th percentile (boxes). Data from four independent experiments.



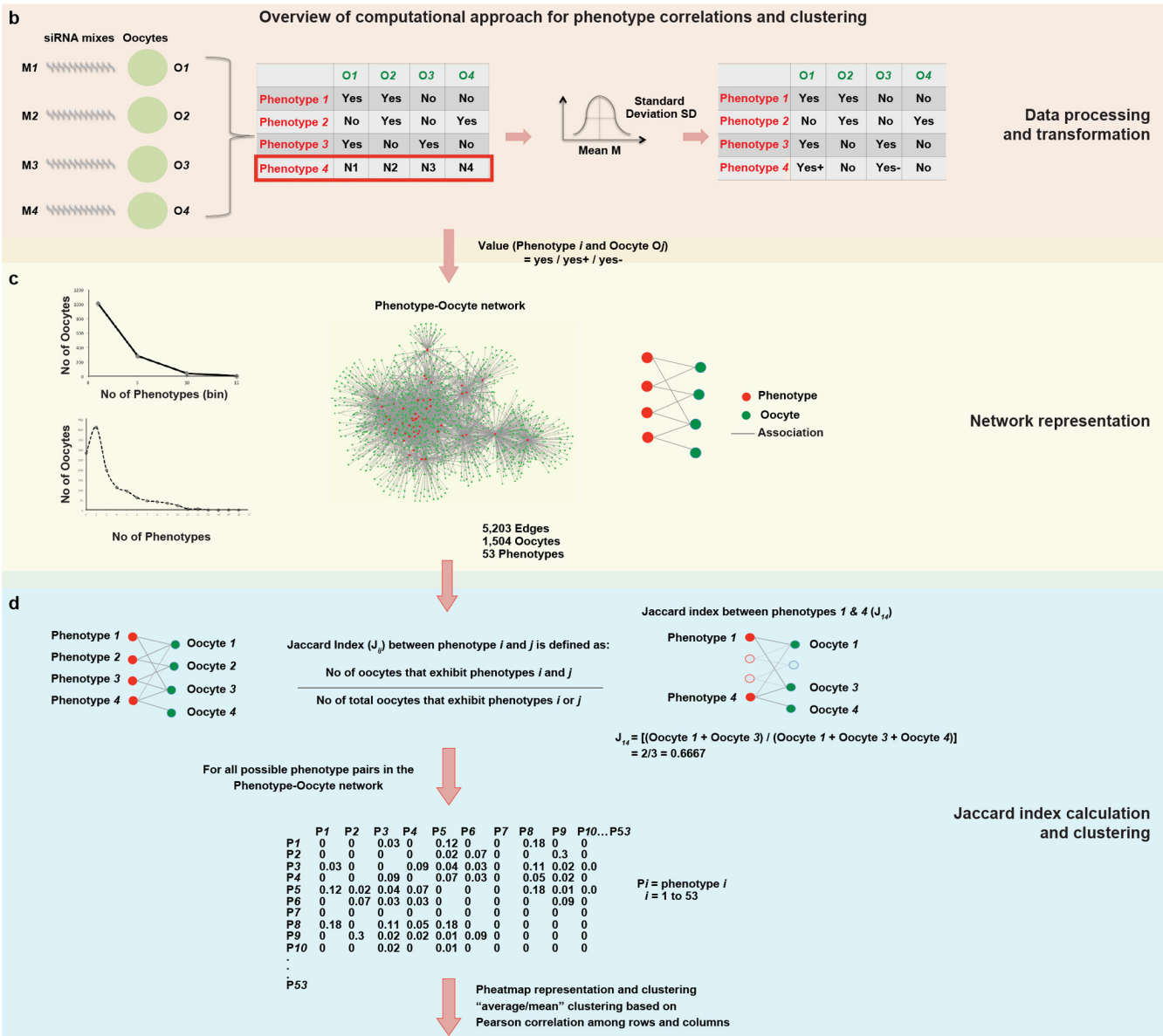
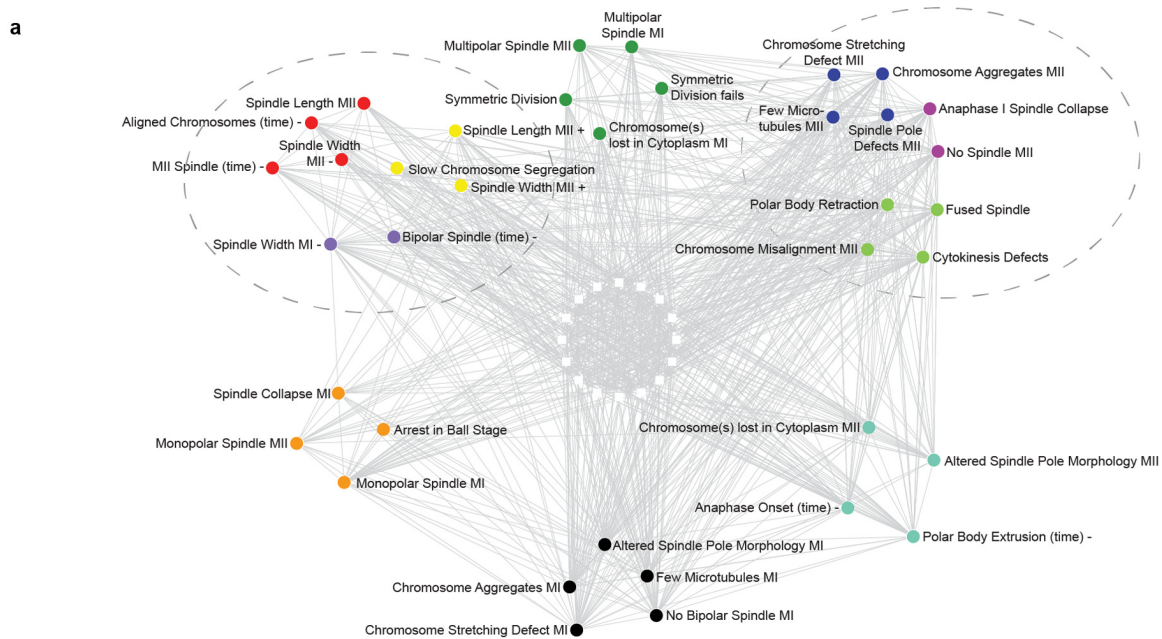
Extended Data Figure 8 | *Mastl* is required for metaphase II arrest and accurate chromosome segregation, but is dispensable for cytokinesis and chromosome condensation in mouse oocytes. **a**, Live oocytes microinjected with control siRNA (control) or siRNAs targeting *Mastl* (*Mastl* RNAi) expressing eGFP–Lamin B1 (green, nuclear lamina) and H2B–mRFP (magenta, chromosomes) merged with DIC. Representative of 30 control and 16 *Mastl* RNAi oocytes. Scale bar, 10 μ m. **b, c**, Live oocytes microinjected with a mix of three different *Mastl* siRNAs expressing human Greatwall fused with eGFP (green) and H2B–mRFP (magenta, chromosomes) merged with DIC. eGFP–Greatwall localized to the nucleus and was released into the cytoplasm shortly before NEBD, consistent with previous studies in mitotic cells^{29,30}. Representative of 23 oocytes. Quantification in Fig. 3f. Scale bar, 10 μ m. **d–g**, Live oocytes microinjected with control siRNA or *Mastl* siRNAs were monitored by long-term time-lapse microscopy and scored for anaphase progression (**d**), time of anaphase onset (**e**), successful formation or retraction

of a polar body upon anaphase (**f**) and the prolonged presence of a midbody upon cytokinesis (**g**). The number of analysed oocytes is specified next to bars. Data from five independent experiments. **h**, Maximum z-projection (left) and three-dimensional reconstruction (right) of chromosomes (Hoechst) in fixed mouse oocytes microinjected with control siRNAs or siRNAs targeting *Mastl* were generated in Imaris. Quantification in **i**. **i**, The chromosome volume was quantified in mouse oocytes microinjected with control siRNAs or siRNAs targeting *Mastl* as shown in **h** in Imaris. The number of analysed oocytes is specified next to bars. Data from two independent experiments. **j**, *Mastl* mRNA levels in control oocytes and oocytes microinjected with *Mastl* siRNAs were quantified by real-time PCR. Mean values from two independent experiments. *P* values were calculated with Fisher's exact test (**d, f, g**) or Student's *t*-test (**e, i**). The box plots in **e** and **j** show median (line), mean (small square), 5th, 95th (whiskers) and 25th and 75th percentile (boxes).



Extended Data Figure 9 | Systematic analysis of phenotype correlations in mouse oocytes. **a, b,** Heatmap representation of clusters of phenotypes generated based on Jaccard indices between them. Jaccard indices, range between 0 and 1, were calculated as described in Methods and Extended Data Fig. 10. Jaccard indices calculated from control oocytes (**a**) and RNAi-treated

oocytes (**b**) are shown. The 'red' and 'blue' respectively correspond to high and low Jaccard indices as indicated by the legend. Clusters of phenotypes were generated using Pheatmap with 'Pearson correlation' values and 'average' clustering input parameters.



Extended Data Figure 10 | Network of phenotypes and calculation of Jaccard indices. **a**, Network of phenotype to oocytes was converted into a phenotype–phenotype network based on number of oocytes that display two phenotypes in question. The network consists of 53 phenotypes and 867 connections between them. The nodes in the network denote phenotypes and edges denote shared oocytes. This is a qualitative network and does not consider the strength of connection, edge weight or number of oocytes in which a given pair of phenotypes co-occurs. Nodes of identical colours denote a cluster (a group of related phenotypes based on topological properties of the network). Phenotypes that are not part of any cluster are in the centre and indicated by squares (white). Related clusters (if they share phenotypes) are marked by dashed circles considered as ‘superclusters’. Clusters were identified by the NeMo method in Cytoscape. Network clusters are purely based on topological properties and are in agreement with the clusters in the heatmap constructed using quantitative measures of Jaccard indices (Extended Data Fig. 9a): for example, two superclusters, top left and top right respectively, correspond to heatmap clusters at the top left and middle of Extended Data

Fig. 9a. **b–d**, Overview of computational approach with schematics to decipher phenotype clusters. O_i , M_i and N_i correspond to oocyte i , mix i and numerical value of phenotype i , respectively. **b**, Conversion of yes, no and numerical data depicts the way we converted a combination of ‘yes’, ‘no’ and numerical data (denoted by N_1 , N_2 , N_3 and N_4) of phenotypes across oocytes into purely ‘yes’ and ‘no’ groups with the ‘yes’ group further classified as ‘yes+’ and ‘yes–’. **c**, Reconstruction of the phenotype–oocyte network: we reconstructed a phenotype–oocyte network from the above data of ‘yes’ and ‘no’ values by considering only the ‘yes’ group. A nonlinear decay relationship between the number of phenotypes and number of oocytes in the network is displayed as represented by two plots. Details of the plots suggest a median value of 2 for phenotypes. **d**, Network transformation and calculation of Jaccard index matrix illustrate our network transformation strategy from a phenotype–oocyte network to a phenotype–phenotype network and the simultaneous estimation of Jaccard indices between phenotypes. The matrix of Jaccard indices between phenotypes was clustered using the pheatmap software in the R package with the ‘Pearson correlation’ parameter and the ‘average’ clustering method.

The CREB coactivator CRTC2 controls hepatic lipid metabolism by regulating SREBP1

Jinbo Han^{1*}, Erwei Li^{1*}, Liqun Chen¹, Yuanyuan Zhang¹, Fangchao Wei¹, Jieyuan Liu², Haiteng Deng² & Yiguo Wang¹

Abnormal accumulation of triglycerides in the liver, caused in part by increased *de novo* lipogenesis, results in non-alcoholic fatty liver disease and insulin resistance^{1,2}. Sterol regulatory element-binding protein 1 (SREBP1), an important transcriptional regulator of lipogenesis, is synthesized as an inactive precursor that binds to the endoplasmic reticulum (ER). In response to insulin signalling, SREBP1 is transported from the ER to the Golgi in a COPII-dependent manner, processed by proteases in the Golgi, and then shuttled to the nucleus to induce lipogenic gene expression^{3–5}; however, the mechanisms underlying enhanced SREBP1 activity in insulin-resistant obesity and diabetes remain unclear. Here we show in mice that CREB regulated transcription coactivator 2 (CRTC2)⁶ functions as a mediator of mTOR⁷ signalling to modulate COPII-dependent SREBP1 processing. CRTC2 competes with Sec23A, a subunit of the COPII complex⁸, to interact with Sec31A, another COPII subunit, thus disrupting SREBP1 transport. During feeding, mTOR phosphorylates CRTC2 and attenuates its inhibitory effect on COPII-dependent SREBP1 maturation. As hepatic overexpression of an mTOR-defective CRTC2 mutant in obese mice improved the lipogenic program and insulin sensitivity, these results demonstrate how the transcriptional coactivator CRTC2 regulates mTOR-mediated lipid homeostasis in the fed state and in obesity.

Insulin resistance, which is associated with hyperglycaemia and hypertriglyceridaemia, is the central problem of type 2 diabetes^{1,2,9}. Although the mechanisms that underlie enhanced glucose and triglyceride levels remain elusive, hepatic lipogenesis and gluconeogenesis are well known to contribute to the paradoxical effects of insulin resistance^{1,2,9–11}. Hepatic lipogenesis is regulated in a combinatorial manner by transcription factors including peroxisome proliferator-activated receptor gamma (PPAR γ), liver X receptor (LXR), carbohydrate response element binding protein (ChREBP) and sterol regulatory element-binding protein-1c (SREBP-1c), while gluconeogenesis is regulated by forkhead box protein O1 (FOXO1), PPAR γ coactivator-1 α (PGC-1 α), cAMP response element-binding protein (CREB) and CREB regulated transcription coactivators (CRTC2s) at the transcriptional level^{5,6,10,11}. CRTC2s are extensively studied in glucose homeostasis, whereas previous studies have suggested possible roles of CRTC2s in lipid homeostasis^{12–14}. These findings prompted us to investigate whether CRTC2 directly affects lipid levels in the liver, where CRTC2 is highly expressed.

We measured hepatic lipid levels in *Crtc2*^{+/+} and *Crtc2*^{-/-} mice. Hepatic triglycerides, but not cholesterol, are increased by 50% in *Crtc2*^{-/-} mice compared to *Crtc2*^{+/+} mice fed with a regular diet; the ratio is even higher under high-fat diet (HFD) feeding conditions (Fig. 1a, b and Extended Data Fig. 1a). Transcript analysis revealed specific upregulation of SREBP1-target genes involved in triglyceride synthesis in *Crtc2*^{-/-} mice fed with both regular and high-fat diets (Extended Data Fig. 1b). SREBP1 (also called SREBF1) belongs to the family of basic helix–loop–helix-leucine zipper transcription factors

synthesized as inactive precursors bound to the ER^{4,5}. Upon sensing insulin stimulation or sterol depletion, SREBP1 is transported to the Golgi through COPII-mediated vesicle trafficking, released by a two-step proteolytic cleavage, and then shuttled to the nucleus to induce the expression of genes involved in cholesterol and fatty acid synthesis^{3–5}. On the basis of these results, we checked whether CRTC2 modulates nuclear SREBP protein levels. In *Crtc2*^{-/-} mice fed with both regular and high-fat diets, the active, nuclear-localized SREBP1 (nSREBP1) was significantly increased while the full-length SREBP1 (fSREBP1) was slightly decreased (Fig. 1c), although total *Srebp1c* mRNA levels remained unchanged (Extended Data Fig. 1b), suggesting that CRTC2 mediates SREBP1 activity at the post-transcriptional level. By contrast, *Crtc2* deficiency decreased insulin levels and did not modulate the activity of SREBP2, a master regulator of cholesterol synthesis⁵, which is consistent with the lack of cholesterol accumulation (Extended Data Fig. 1a). In addition, the abnormal accumulation of nSREBP1 and the enhanced lipogenic profile in *Crtc2*^{-/-} mice were normalized to the level found in fed *Crtc2*^{+/+} mice by adenovirus-mediated wild-type (WT) CRTC2 overexpression or knockdown of *Srebp1* (Fig. 1d and Extended Data Fig. 2). Taken together, these results suggest that CRTC2 modulates triglyceride synthesis via regulation of SREBP1 maturation.

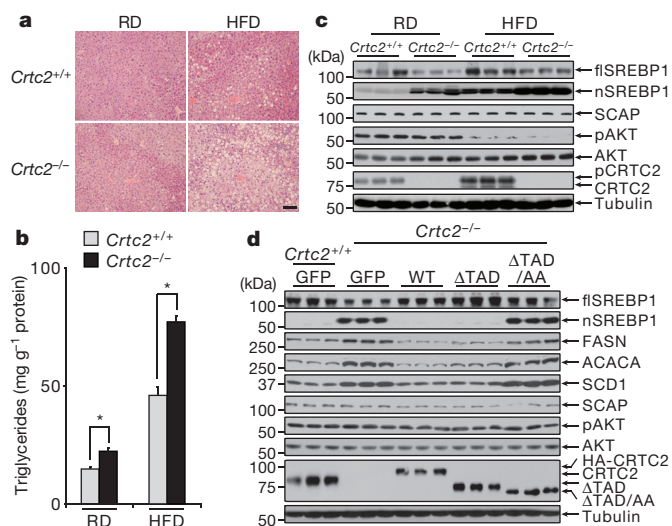


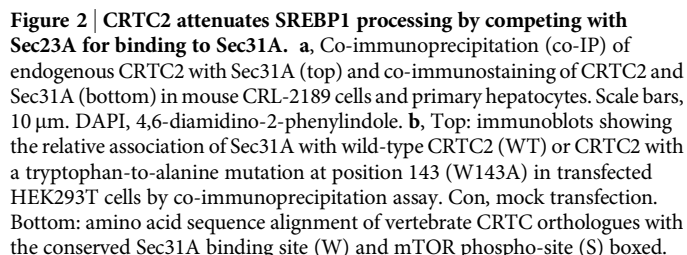
Figure 1 | Accumulation of hepatic lipids and mature SREBP1 in *Crtc2*^{-/-} mice. **a–c**, Haematoxylin and eosin sections of liver (**a**), levels of hepatic triglycerides (**b**) and immunoblots showing hepatic amounts of full-length, inactive SREBP1 (fSREBP1) and cleaved, active SREBP1 (nSREBP1) in liver extracts (**c**) from *Crtc2*^{+/+} and *Crtc2*^{-/-} mice fed with a regular diet (RD) or high-fat diet (HFD) for 18 weeks. Scale bar, 50 μ m. Data are shown as mean \pm s.e.m. * $P < 0.01$, $n = 10$. **d**, Effect of CRTC2 and its mutants (CRTC2(Δ TAD), amino acids 1–630; CRTC2(Δ TAD/AA), amino acids 1–630 with double alanine mutations at serine 171 and serine 275) on SREBP1 maturation and lipogenic protein levels (FASN, SCD1 and ACACA).

¹MOE Key Laboratory of Bioinformatics, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China. ²Proteomics Facility, School of Life Sciences, Tsinghua University, Beijing 100084, China.

*These authors contributed equally to this work.

On the basis of our previous results that CRTC2 is peripherally associated with the ER^{15,16}, we hypothesized that CRTC2 may bind to proteins involved in the regulation of SREBP1 transport. Indeed, CRTC2 binds to Sec31A, a subunit of the COPII complex⁸, as identified by mass spectrometry and confirmed by co-immunoprecipitation and co-immunostaining assays (Fig. 2a and Extended Data Fig. 4a–d). Further analysis showed that Trp143 of CRTC2 is important for the CRTC2–Sec31A interaction, since a tryptophan-to-alanine mutant (W143A) abolished the interaction between CRTC2 and Sec31A, without perturbing its subcellular localization and transcriptional activity (Fig. 2b and Extended Data Fig. 4e–g). Conversely, CRTC2 associates with a carboxy-terminal polypeptide of Sec31A (Fig. 2c). Since the C terminus of Sec31A also mediates its interaction with Sec23A (ref. 17), we proposed that CRTC2 may block Sec23A–Sec31A binding. Supporting this notion, increased amounts of CRTC2 attenuate Sec23A binding to Sec31A, and vice versa (Extended Data Fig. 4h). However, Sec23A, with higher affinity to Sec31A, was more potent at disrupting the CRTC2–Sec31A interaction.

It is well established that the processing and nuclear activity of SREBP1 is induced in response to insulin signalling and nutrient signalling^{3,5,18}. We hypothesized that the CRTC2–Sec31 interaction may modulate hormonal and nutritional activation of SREBP1. In fact, both insulin and amino acid stimulation attenuated the CRTC2–Sec31A interaction with concomitant enhancement of the Sec23A–Sec31A association (Extended Data Fig. 6a, b). As shown in Extended Data Fig. 6a, b, the regulatory effect of CRTC2 on the Sec23–Sec31A interaction was abolished in the presence of torin1, an inhibitor of mTOR that controls lipid metabolism and SREBP1 activation^{7,18–21}. To examine whether mTOR directly phosphorylates CRTC2, Sec23A or Sec31A and modulates the association of this complex, we identified their phosphorylation sites by mass spectrometry analysis (Extended Data Fig. 6c and data not shown for Sec23A, Sec31A). The conserved serine site at position 136 (Ser136) of CRTC2, which occurs in the context of a classic mTOR substrate motif (S/T-P)²², was phosphorylated without rapamycin treatment and dephosphorylated in the presence of rapamycin (Figs 2b and 3a and Extended Data



c, Deletion analysis of regions in Sec31A required for the CRTC2–Sec31A interaction. Interaction-competent Sec31A polypeptides are indicated by (+) in each schematic. **d**, *In vitro* pull-down assay showing the binding ability of His-tagged full-length CRTC2, CRTC2(W143A) mutant and Sec23A to the C terminus of His/HA–Sec31A (amino acids 701–1220). The asterisk shows an unspecific protein band. **e**, **f**, Effect of wild-type (WT) and Sec31A-binding-defective (W143A) CRTC2 on HA-tagged SREBP1c transport from the ER to the Golgi in an *in vitro* budding assay (**e**) and SREBP1 processing in fed mice (**f**). The constitutive budding protein (ERGIC-53) and ER lumen proteins (GRP94 and GRP78) are indicated as controls (**e**).

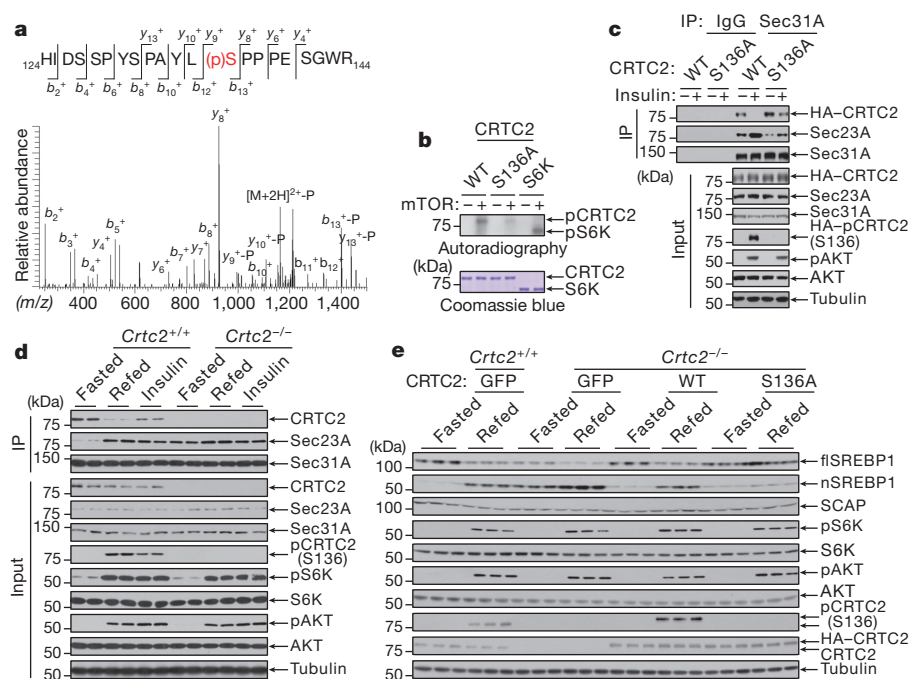


Figure 3 | mTOR phosphorylates CRTC2 and promotes COPII-dependent SREBP1 activation. **a**, Identification of CRTC2 Ser136 phosphorylation by liquid chromatography mass spectrometry (LC-MS/MS) analysis. The MS/MS spectrum of the phosphorylated peptide HIDSSPYSPAYL(p) SPPPESGWR is shown. The labelled peaks show the masses of *y* or *b* ions of the phosphorylated peptide. **b**, *In vitro* kinase assay showing the phosphorylation of His-tagged CRTC2 and S6K by truncated mTOR. **c**, Immunoblots of co-immunoprecipitation assay showing the effect of wild-type and the mTOR-defective mutant CRTC2(S136A) on Sec31A–Sec23A interaction in mouse primary hepatocytes in response to 100 nM insulin stimulation for 30 min. **d**, Regulatory effect of CRTC2 on Sec31A–Sec23A binding by fasting, refeeding or insulin treatment in *Crtc2*^{+/+} and *Crtc2*^{−/−} mice. Mice were fasted for 3 h then refed for 1 h or treated with intraperitoneal administration of insulin (0.1 U kg^{−1}) for 1 h. **e**, Effect of wild type and the S136A mutant of CRTC2 on SREBP1 maturation in fasted (3 h) and refed (1 h after 3 h fasting) mice.

Fig. 6c). In addition, a co-immunoprecipitation assay revealed that mTOR interacts with CRTC2, and an *in vitro* kinase assay showed that mTOR directly phosphorylates CRTC2 at Ser136 (Fig. 3b and Extended Data Fig. 6d). To confirm the phosphorylation of CRTC2 at Ser136 *in vivo*, we raised a polyclonal antibody against the phospho-CRTC2 (Ser136) peptide and found that both insulin and amino acids stimulate CRTC2 Ser136 phosphorylation in an mTOR-dependent manner, suggesting that CRTC2 is a bona fide substrate of mTOR (Fig. 3c and Extended Data Fig. 6a, b). Phosphorylation of CRTC2 Ser136 is much more sensitive to torin1 than to rapamycin treatment (Extended Data Fig. 6e). Although mTOR-defective CRTC2(S136A) had similar cellular localization and nuclear activity on Cre-luc to wild-type CRTC2 (Extended Data Fig. 6f, g), it reduced the insulin-stimulated enhancement of Sec23A–Sec31A association (Fig. 3c).

Since CRTC2 was able to be phosphorylated in primary hepatocytes by mTOR, we next checked whether CRTC2 could be phosphorylated via fasting–refeeding transition in mice, thereby modulating the Sec23A–Sec31A interaction and SREBP1 activation. Refeeding led to CRTC2 phosphorylation and disrupted its association with Sec31A, which became available for Sec23A binding, thus enhancing the Sec23A–Sec31A interaction; insulin had a similar but weaker effect (Fig. 3d). However, the dynamic regulation of Sec31A–Sec23A by CRTC2 was lost in *Crtc2*^{−/−} mice (Fig. 3d). In addition, wild-type CRTC2 normalized nSREBP1 in *Crtc2*^{−/−} mice during both fasting and refeeding to the level in *Crtc2*^{+/+} mice, and was phosphorylated during refeeding (Fig. 3e). The CRTC2(S136A) mutant had a stronger inhibitory effect on SREBP1 processing and lipogenic gene expression during refeeding owing to deficient phosphorylation by mTOR (Fig. 3e and Extended Data Fig. 7a, b). Furthermore, CRTC2(S136A) did not affect the outcome of torin1 treatment on SREBP1 activation, hepatic triglyceride levels, lipin1 expression or cellular localization (Extended Data Fig. 7c–e), suggesting that both CRTC2 and lipin1, as mTOR downstream mediators, regulate SREBP1 activation at different steps¹⁸. Taken together, these results demonstrate that mTOR modulates COPII-exposed SREBP1 processing via Ser136 phosphorylation of CRTC2.

Considering that lipogenesis is chronically enhanced in obesity and diabetes^{1,2,5,11}, we tested whether the mTOR-CRTC2 axis is altered in this setting. Consistent with previous studies^{23–25}, hepatic nuclear SREBP1, mTOR activity, and levels of branched-chain amino acids

and triglycerides were all increased in HFD-fed, *db/db* and *ob/ob* mice (Fig. 4a and Extended Data Fig. 8). Meanwhile, the Sec31A–Sec23A interaction and the level of phospho-CRTC2(Ser136) were enhanced after HFD feeding. Accordingly, the association of CRTC2–Sec31A was blocked, although the protein level of COPII subunits remained stable (Fig. 4a and Extended Data Fig. 8c). Since the CRTC2(S136A) mutant blocks SREBP1 processing, we asked whether this mTOR-defective mutant was able to reduce hepatic hypertriglyceride levels by inhibiting SREBP1 activation in obesity. To exclude a gluconeogenic effect of CRTC2(S136A), we used an mTOR-defective CRTC2 mutant without the transactivation domain (CRTC2(Δ TAD/S136A)) to restore SREBP1 processing in HFD-fed mice. CRTC2(Δ TAD/S136A) did not affect mouse body weight, fat mass, food intake, energy expenditure and liver function measured by alanine aminotransferase (ALT) and aspartate aminotransferase (AST) activity, but it reduced SREBP1 processing, the lipogenesis rate and gluconeogenic capacity, as well as improving hepatic steatosis and insulin sensitivity (Fig. 4b–e and Extended Data Fig. 9a–f). To investigate further the role of CRTC2(Δ TAD/S136A) in insulin sensitivity, we performed hyperinsulinaemic-euglycaemic clamp studies. The steady-state glucose infusion rate almost doubled during the clamp studies in CRTC2(Δ TAD/S136A) mice, reflecting enhanced whole-body insulin responsiveness, and was accompanied by an increase in the glucose disposal rate (Fig. 4f, g). The insulin-stimulated glucose disposal rate, which primarily reflects skeletal muscle insulin sensitivity, and the insulin-induced suppression of plasma free fatty acid levels, which indicates white adipose tissue insulin sensitivity, were both slightly increased in the presence of hepatic CRTC2(Δ TAD/S136A), suggesting a possible role of inter-organ communication in orchestrating systemic insulin sensitivity. However, the insulin sensitivity in liver was markedly improved as judged from insulin-induced suppression of hepatic glucose production and further confirmed by pAKT levels (Fig. 4h and Extended Data Fig. 9g–k).

Previous studies showed that nuclear CRTC2, as a transcriptional coactivator, plays an important part in gluconeogenesis during fasting or ER stress via its binding to different partners⁶. Here, our results demonstrate that cytosolic CRTC2, as a critical mediator of mTOR, modulates COPII activity, which leads to SREBP1 processing and enhancement of *de novo* lipogenesis (Fig. 4i), thereby contributing to hepatic lipid levels and insulin resistance along with potential alterations in free fatty acid supply from either dietary or adipose tissue

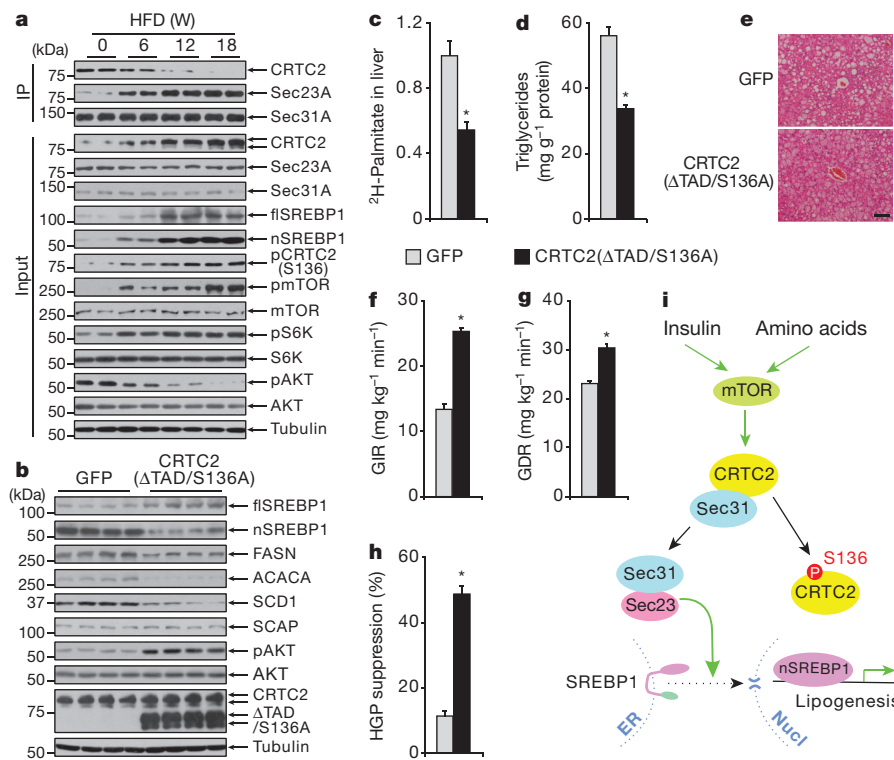


Figure 4 | Dysregulated SREBP1 signalling in obese mice is improved by overexpression of mTOR-defective CRTC2. **a**, Immunoblots showing co-immunoprecipitation of CRTC2 and Sec23A with Sec31A in livers of mice fed with a HFD for different times (W, weeks). **b–e**, Effect of the mTOR-defective mutant CRTC2(ΔTAD/S136A) on SREBP1 maturation (**b**), relative synthesis rate of hepatic palmitate (**c**), levels of hepatic triglycerides (**d**), and hepatic steatosis shown by haematoxylin and eosin staining of liver (**e**) in mice fed with a HFD for 18 weeks. **f–h**, Glucose infusion rate (GIR; **f**), glucose disposal rate (GDR; **g**) and percentage suppression of hepatic glucose production (HGP; **h**) during hyperinsulinaemic-euglycaemic clamp studies. **i**, mTOR modulates COPII-dependent SREBP1 activation via Ser136 phosphorylation of CRTC2. During feeding, Sec31A dissociates from CRTC2, which becomes phosphorylated at Ser136 by mTOR, and interacts instead with Sec23A, thus promoting COPII-dependent transport and processing of SREBP1 in the Golgi. In obesity, enhanced phosphorylation of CRTC2 by mTOR contributes at least in part to increased SREBP1 activation and hepatic lipogenesis. Scale bar, 50 μm. Data are shown as mean ± s.e.m. **P* < 0.01, *n* = 8 (**c**, **d**), *n* = 6 (**f–h**).

lipolysis^{1,2}. Our findings expand the role of the transcription coactivator CRTC2 to include lipid metabolism, and provide insight into how SREBP1 activity is enhanced in obesity and diabetes. Considering the isoform diversity of COPII subunits and CRTCs^{6,8}, as well as the environmental cues that activate mTOR⁷, mTOR–CRTC signalling may have important functions in regulating other cargo transport processes in different settings. Therefore, it will be interesting to explore the selectivity and specificity of this signalling axis in the future.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 June 2014; accepted 28 April 2015.

Published online 6 July 2015.

- Samuel, V. T. & Shulman, G. I. Mechanisms for insulin resistance: common threads and missing links. *Cell* **148**, 852–871 (2012).
- Perry, R. J., Samuel, V. T., Petersen, K. F. & Shulman, G. I. The role of hepatic lipids in hepatic insulin resistance and type 2 diabetes. *Nature* **510**, 84–91 (2014).
- Yellaturu, C. R. *et al.* Insulin enhances post-translational processing of nascent SREBP-1c by promoting its phosphorylation and association with COPII vesicles. *J. Biol. Chem.* **284**, 7518–7532 (2009).
- Goldstein, J. L., DeBose-Boyd, R. A. & Brown, M. S. Protein sensors for membrane sterols. *Cell* **124**, 35–46 (2006).
- Horton, J. D., Goldstein, J. L. & Brown, M. S. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J. Clin. Invest.* **109**, 1125–1131 (2002).
- Altarejos, J. Y. & Montminy, M. CREB and the CRTC co-activators: sensors for hormonal and metabolic signals. *Nature Rev. Mol. Cell Biol.* **12**, 141–151 (2011).
- Lamming, D. W. & Sabatini, D. M. A Central role for mTOR in lipid homeostasis. *Cell Metab.* **18**, 465–469 (2013).
- Zanetti, G., Pahuja, K. B., Studer, S., Shim, S. & Schekman, R. COPII and the regulation of protein sorting in mammals. *Nature Cell Biol.* **14**, 20–28 (2012).
- Brown, M. S. & Goldstein, J. L. Selective versus total insulin resistance: a pathogenic paradox. *Cell Metab.* **7**, 95–96 (2008).
- Lin, H. V. & Accili, D. Hormonal regulation of hepatic glucose production in health and disease. *Cell Metab.* **14**, 9–19 (2011).
- Strable, M. S. & Ntambi, J. M. Genetic control of *de novo* lipogenesis: role in diet-induced obesity. *Crit. Rev. Biochem. Mol. Biol.* **45**, 199–214 (2010).
- Wang, Y. *et al.* Targeted disruption of the CREB coactivator Crtc2 increases insulin sensitivity. *Proc. Natl Acad. Sci. USA* **107**, 3087–3092 (2010).
- Song, Y. *et al.* CRTC3 links catecholamine signalling to energy balance. *Nature* **468**, 933–939 (2010).

- Altarejos, J. Y. *et al.* The Creb1 coactivator Crtc1 is required for energy balance and fertility. *Nature Med.* **14**, 1112–1117 (2008).
- Wang, Y. *et al.* Inositol-1,4,5-trisphosphate receptor regulates hepatic gluconeogenesis in fasting and diabetes. *Nature* **485**, 128–132 (2012).
- Wang, Y., Vera, L., Fischer, W. H. & Montminy, M. The CREB coactivator CRTC2 links hepatic ER stress and fasting gluconeogenesis. *Nature* **460**, 534–537 (2009).
- Bi, X., Mancias, J. D. & Goldberg, J. Insights into COPII coat nucleation from the structure of Sec23•Sar1 complexed with the active fragment of Sec31. *Dev. Cell* **13**, 635–645 (2007).
- Peterson, T. R. *et al.* mTOR complex 1 regulates lipin 1 localization to control the SREBP pathway. *Cell* **146**, 408–420 (2011).
- Li, S., Brown, M. S. & Goldstein, J. L. Bifurcation of insulin signaling pathway in rat liver: mTORC1 required for stimulation of lipogenesis, but not inhibition of gluconeogenesis. *Proc. Natl Acad. Sci. USA* **107**, 3441–3446 (2010).
- Porstmann, T. *et al.* SREBP activity is regulated by mTORC1 and contributes to Akt-dependent cell growth. *Cell Metab.* **8**, 224–236 (2008).
- Yecies, J. L. *et al.* Akt stimulates hepatic SREBP1c and lipogenesis through parallel mTORC1-dependent and independent pathways. *Cell Metab.* **14**, 21–32 (2011).
- Hsu, P. P. *et al.* The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* **332**, 1317–1322 (2011).
- Khamzina, L., Veilleux, A., Bergeron, S. & Marette, A. Increased activation of the mammalian target of rapamycin pathway in liver and skeletal muscle of obese rats: possible involvement in obesity-linked insulin resistance. *Endocrinology* **146**, 1473–1481 (2005).
- Um, S. H. *et al.* Absence of S6K1 protects against age- and diet-induced obesity while enhancing insulin sensitivity. *Nature* **431**, 200–205 (2004).
- Shimomura, I. *et al.* Decreased IRS-2 and increased SREBP-1c lead to mixed insulin resistance and sensitivity in livers of lipodystrophic and ob/ob mice. *Mol. Cell* **6**, 77–86 (2000).

Acknowledgements We thank M. Montminy, W. E. Balch, L. Yu, S. Fu, P. Li, J. Wu and D. Li for discussion, and X. Liu and P. Li for technical help with lipogenesis analyses and clamp studies, respectively. This work was supported by grants from 1000 Talents Program for Young Scholars and the National Science Foundation of China (31322027, 31321003 and 31471127) to Y.W.

Author Contributions J.H., E.L. and Y.W. designed the study and analysed the data. J.H., E.L., L.C., Y.Z., F.W. and Y.W. performed the experiments. J.L. and H.D. carried out mass spectrometry analysis. J.H. and Y.W. wrote the paper. All authors reviewed and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.W. (wangyiguo@biomed.tsinghua.edu.cn).

METHODS

Mouse strains. Mice were housed in colony cages with a 12 h light/dark cycle in a temperature-controlled environment. For high-fat diet feeding experiments, regular diet (Research Diets, D12450B) was replaced with a diet containing 60 kcal% fat (Research Diets, D12492). *Crtc2*^{-/-} mice have been described previously¹². Heterozygous *Crtc2*^{+/-} mice were backcrossed with C57BL/6J for ten generations and then intercrossed to obtain *Crtc2*^{-/-} mice confirmed by PCR-based genotyping. All the other mice were purchased from Jackson Labs (Bar Harbour, ME). All animal experiments were approved by the Animal Care and Use Committee at Tsinghua University.

Plasmids and adenoviruses. Plasmids containing human Sec13, Sec31A and Sec23A were provided by W. E. Balch (The Scripps Research Institute). HA-tagged S6K (8984) and Myc-tagged mTOR (1861) plasmids were from Addgene. Myc-DDK-tagged lipin1 (RC207138) plasmid was purchased from OriGene. Adenoviruses (1 × 10⁸ plaque forming units (pfu) GFP, CRTC2, CRTC2(W143A), CRTC2(S136A), CRTC2(ΔTAD) (amino acids 1–630), CRTC2(ΔTAD/AA) (S171A, S275A), CRTC2(ΔTAD/S136A), *Srebp1* RNAi, or unspecific RNAi) were delivered to male C57BL/6J, HFD-fed, *Crtc2*^{+/-}, or *Crtc2*^{-/-} mice by tail vein injection. Mice were injected with adenovirus on day 0 and killed on day 7. Wild-type CRTC2, *Crtc2* RNAi, GFP and unspecific RNAi adenoviruses have been described previously^{15,16}. CRTC2(W143A), CRTC2(S136A), CRTC2(ΔTAD), CRTC2(ΔTAD/AA) and CRTC2(ΔTAD/S136A) adenoviruses were constructed from mouse *Crtc2* cDNA. *Srebp1* RNAi adenovirus was constructed using the sequence 5'-GGGATCAAAGAGGAGCCAGTGC-3'. All expressed constructs used in this study were confirmed by sequencing.

In vivo analysis and histology. Triglyceride (Sigma, TR0100) and cholesterol (BioVision, K603-100) levels in liver and plasma, plasma insulin (Mercodia, 10-1247-01), plasma alanine aminotransferase (ALT, BioVision, K752-100), plasma aspartate aminotransferase (AST, BioVision, K753-100), hepatic branched-chain amino acids (BCAA, Sigma, MAK003) and plasma free fatty acid levels (BioVision, K612-100) were measured according to the manufacturer's instructions. Blood glucose values were determined using a LifeScan automatic glucometer. Insulin tolerance tests, glucose tolerance tests and pyruvate tolerance tests were conducted as previously reported¹². *De novo* lipogenesis experiments were performed as previously reported²⁶. Pieces of liver tissue were sent to the Metabolomics Center at Tsinghua University to determine the ²H₂O incorporated palmitate levels by liquid chromatography and mass spectrometry. Hyperinsulinaemic-euglycaemic clamps were performed as previously described²⁷. Three days after adenovirus administration, mice were implanted with dual jugular catheters for 4 days for use in clamp studies. Food intake and energy expenditure were simultaneously measured for individually housed mice with a PhenoMaster system (TSE Systems). Relative fat mass was measured with an EchoMRI analyser. For histology, mouse tissues were fixed in 4% paraformaldehyde and paraffin embedded. Sections (8 μm) were used for haematoxylin and eosin staining.

Quantitative PCR. Total RNA from mouse liver was extracted using RNeasy kits (Qiagen). cDNA was obtained by the iScript cDNA synthesis kit (Bio-Rad). RNA levels were measured with the LightCycler 480 II (Roche) as previously described^{15,16}. The following primers were used for qPCR: *Acaca*-forward GGATGACAGGCTTGACGCTAT, *Acaca*-reverse TTTGTGCAACTAGGAACGTAAGTCG; *Acox1*-forward CTGCCAAGGGACTCCAGAGCAGCT, *Acox1*-reverse GACATGGACACATCCACCTGCAG; *Actin*-forward GTCCACCCCGGGAAGGTGA, *Actin*-reverse AGGCCTCAGACCTGGGCCATT; *Apoa4*-forward GCCCCATGCCAACAAGTAA, *Apoa4*-reverse CCTGTATCGTGGTCTGCATG; *Chrebp1*-forward CTCAGGGTATGCAACCCAGGTG, *Chrebp1*-reverse GACAGGGGTTGTGTCTCTGGC; *Fasn*-forward TGGTGGTGTGGACATGGTCACAGA, *Fasn*-reverse CCGAAGCTGGGGGTCCATTGTGTG; *Gpat1*-forward GCCCTTCGTGGGAAGGTGCTGCTA, *Gpat1*-reverse CCGTCTCGCCAGCCATCCTCTGTG; *Mttp*-forward GAGCGGCTATACAAGCTCAGTAC, *Mttp*-reverse TCACCATCAGGATTCCTCCACAGT; *PPARα*-forward TCTCCACGTTCCAGCCCTTCCTCA, *PPARα*-reverse TTCACATGCGTGAACCTCGTAGTG; *PPARγ*-forward TCCGTGATGGAAGCACTCCGAT, *PPARγ*-reverse CAGCAACCATTGGGTGAGCTCTTG; *Scd1*-forward CTGTA CGGGATCATACTGGTTCCTC, *Scd1*-reverse CAGCCGAGCCTTGTAAGTTCGTG; *Srebp1c*-forward GGAGCCATGGATTGCACAT, *Srebp1c*-reverse GGCCCAGGAGTCACTGT; *Srebp2*-forward GATGAGCTGACTCTCGGGGACCT, *Srebp2*-reverse GTGGGGTAGGAGAGACTTGACCT.

Cell culture and luciferase assay. CRL-2189 and HEK293T (ATCC) cells were cultured in DMEM containing 10% FBS (HyClone), 100 mg ml⁻¹ penicillin-streptomycin. Mouse primary hepatocytes were isolated and cultured as previously described^{15,16}. For reporter studies, Ad-Cre-luc-infected hepatocytes (1 pfu per

cell) or Cre-luc-transfected HEK293T cells were exposed to forskolin (10 μM) for 4 h. All cell lines were routinely tested for mycoplasma presence using a PCR detection kit (Sigma, MP0035).

Immunoblot, immunoprecipitation and immunostaining. Assays were performed as described^{15,16}. CRTC2, pCREB, CREB, AKT, pAKT (Thr308), tubulin, HA and Flag antibodies were as previously described^{15,16}. Other antibodies were purchased as follows: rabbit polyclonal anti-Sec31A (A302-336A), Bethyl Laboratories; mouse monoclonal anti-Sec31A (612351), BD Biosciences; anti-KDEL (ab12223) and anti-Sec12 (ab3422), Abcam; anti-ERGIC-53 (GTX63674), GeneTex; anti-His (D291-3), MBL International; anti-Sar1 (07-692), Millipore; anti-SREBP1 (SC-13551, SC-367), anti-Myc (SC-40), Santa Cruz; anti-SCAP (13102), anti-ACACA (3662), anti-FASN (3180), anti-SCD1 (2438), anti-Sec23A (8162), anti-S6K (2708), anti-pS6K (9234), anti-mTOR (2983) and anti-pmTOR (5536), anti-lipin1 (5195), Cell Signaling Technology; anti-Sec24A (15958-1-AP) and anti-SREBP2 (14508-1-AP), Proteintech Group Inc.; anti-Sec13 (NBP2-20278), Novus Biologicals. The phospho-S136 CRTC2 antibody was generated and purified by Beijing Prorevo Biotech Co., Ltd.

In vitro budding assay. Mouse liver cytosol was harvested as described²⁸. In brief, mouse livers were perfused with 0.9% (w/v) NaCl through the portal vein at room temperature. The perfused livers were disrupted in ice-cold buffer (20 mM HEPES pH 7.4, 150 mM KOAc, 5 mM MgOAc, 250 mM sorbitol) followed by 10 strokes in a 50 ml Dounce homogenizer. Homogenates were centrifuged at 1,000g for 10 min. Supernatants were sequentially centrifuged at 20,000g for 20 min, 186,000g for 1 h, and 186,000g for 45 min. The final supernatant was dialysed, divided into multiple aliquots and stored at -80 °C.

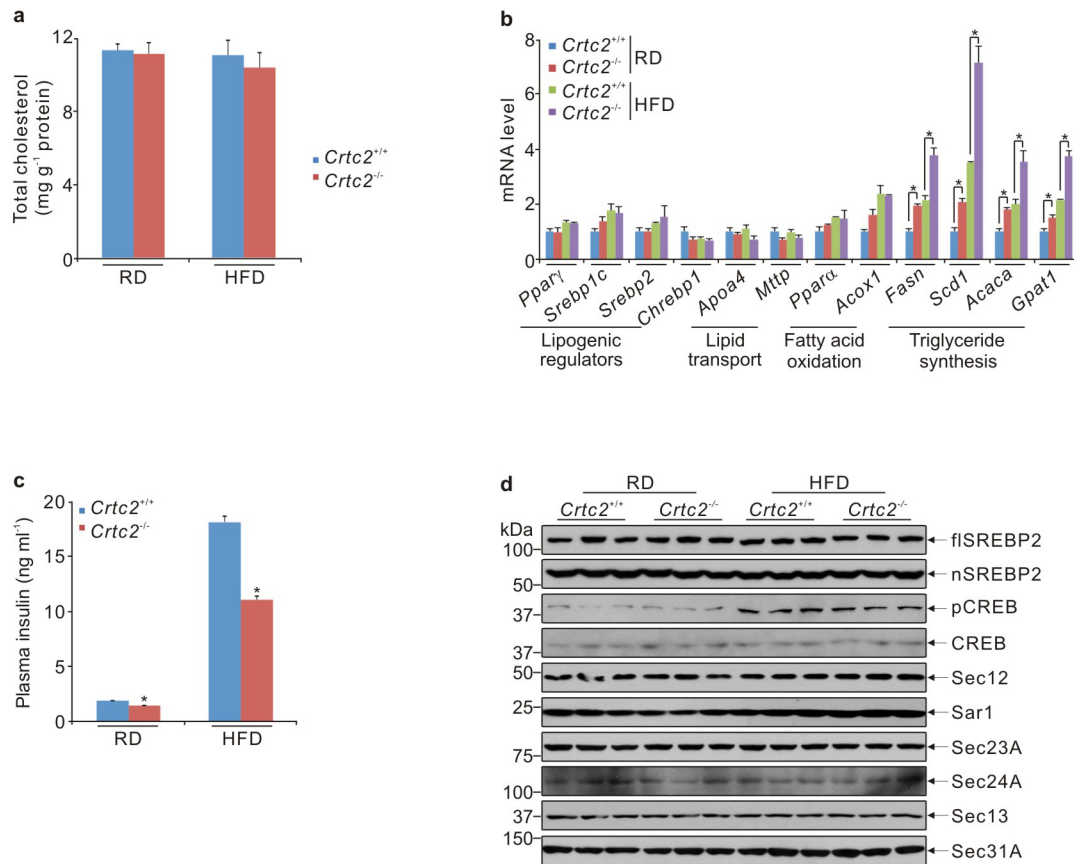
In vitro budding assays were carried out as reported^{28,29}. HEK293T cells were co-transfected with HA-SREBP1c and Flag-SCAP (provided by P. Li, Tsinghua University) for 48 h. Cells were then cultured with lipid-deficient medium for another 3 h, harvested and permeabilized by 5 min incubation in buffer (20 mM HEPES pH 7.4, 150 mM KOAc, 5 mM MgOAc, 250 mM sorbitol) with 40 μg ml⁻¹ digitonin. The permeabilized cells were washed with the same buffer and used at 40 μg protein per reaction. The budding reaction (4 mg ml⁻¹ mouse liver cytosol, 1 mM ATP, 40 mM creatine phosphate, 0.2 mg ml⁻¹ creatine phosphokinase, 0.1 mM GTP, and 100 ng His-CRTC2 or His-CRTC2 (W143A) fusion proteins purified from *Escherichia coli*) was assembled on ice, incubated for 30 min at 30 °C and then put on ice. The donor membranes were removed by centrifugation at 12,000g for 20 min at 4 °C. The supernatant fraction was centrifuged at 4 °C by 25 min 55,000 rpm using a TLA100 rotor in a Beckman Optima TLX ultracentrifuge. The collected vesicles were analysed by SDS-PAGE.

In vitro kinase assay. His-tagged CRTC2, CRTC2(S136A) and S6K fusion proteins were purified from *E. coli*. The kinase assay was performed as reported¹⁸. The reaction system (20 μl), containing 150 ng fusion protein, 20 ng truncated mTOR (Millipore, 14-770) in reaction buffer (25 mM HEPES pH 7.4, 50 mM KCl, 5 mM MgCl₂, and 5 mM MnCl₂), 50 μM cold ATP and 2 μCi [γ -³²P]ATP, was incubated for 30 min at 30 °C. Reactions were stopped by adding 5 μl sample buffer, then boiled for 10 min and analysed by SDS-PAGE followed by autoradiography.

Mass spectrometry (MS). To identify CRTC2-interacting proteins, the ER fraction of CRL-2189 cells was isolated according to the manufacturer's instructions (Sigma-Aldrich, ER0100). Immunoprecipitates of endogenous CRTC2 from the ER fraction were prepared for MS studies as previously reported^{15,16} and analysed by electrospray ionization tandem MS on a Thermo LTQ Orbitrap instrument. To identify the mTOR phospho-site(s) in CRTC2, HEK293T cells were transfected by Flag-CRTC2 and treated with or without 100 nM rapamycin for 1 h. Immunoprecipitates of Flag-CRTC2 were used to determine phospho-site(s) by MS.

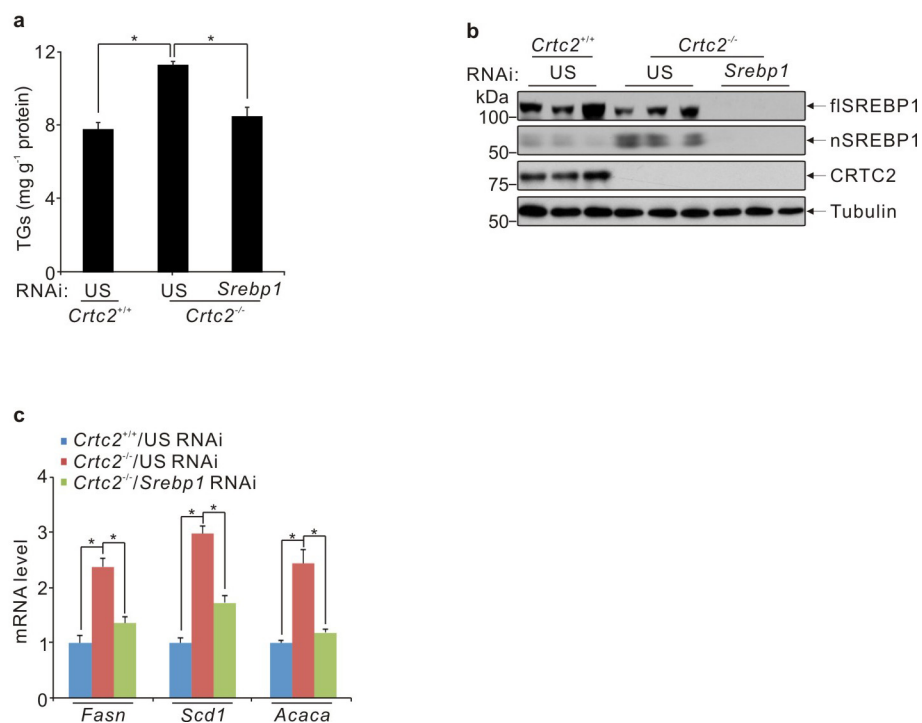
Statistical analyses. Age- and weight-matched male mice were randomly assigned for the experiments. The animal numbers used for all experiments are outlined in the corresponding figure legends. No animals were excluded from statistical analyses, and the investigators were not blinded in the studies. All studies were performed on at least three independent occasions. Results are reported as mean ± s.e.m. Comparison of different groups was carried out using two-tailed unpaired Student's *t*-test. Differences were considered statistically significant at *P* < 0.05. No statistical methods were used to predetermine sample size.

26. Zhao, X. *et al.* Regulation of lipogenesis by cyclin-dependent kinase 8-mediated control of SREBP-1. *J. Clin. Invest.* **122**, 2417–2427 (2012).
27. Li, P. *et al.* Adipocyte NCoR knockout decreases PPAR γ phosphorylation and enhances PPAR activity and insulin sensitivity. *Cell* **147**, 815–826 (2011).
28. Nohturfft, A., Yabe, D., Goldstein, J. L., Brown, M. S. & Espenshade, P. J. Regulated step in cholesterol feedback localized to budding of SCAP from ER membranes. *Cell* **102**, 315–323 (2000).
29. Schindler, A. J. & Schekman, R. *In vitro* reconstitution of ER-stress induced ATF6 transport in COPII vesicles. *Proc. Natl Acad. Sci. USA* **106**, 17775–17780 (2009).

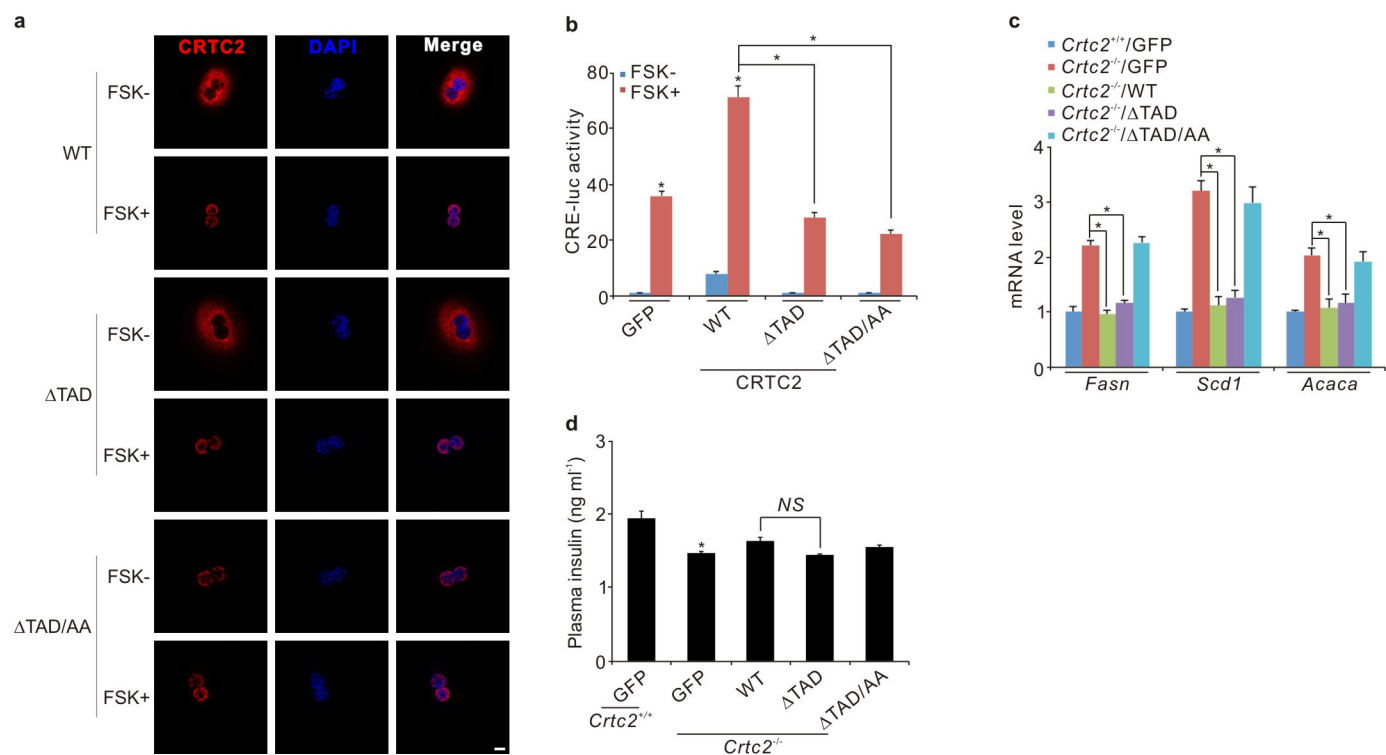


Extended Data Figure 1 | Profiling of cholesterol, gene expression, protein and circulating insulin levels in *Crtc2*^{+/+} and *Crtc2*^{-/-} mice. **a–d**, Hepatic cholesterol levels (**a**), qPCR results for expression of genes involved in lipogenic regulation, lipid transport, fatty acid oxidation and triglyceride synthesis (**b**), plasma insulin level (**c**), and immunoblots showing

hepatic amounts of full-length, inactive SREBP2 (flSREBP2) and cleaved, active SREBP2 (nSREBP2), phospho-CREB (pCREB), CREB, and subunits of the COPII complex (Sec12, Sar1, Sec23A, Sec24A, Sec13 and Sec31A) (**d**) in *Crtc2*^{+/+} and *Crtc2*^{-/-} mice fed with a regular diet (RD) or a high-fat diet (HFD) for 18 weeks. Data are shown as mean ± s.e.m. **P* < 0.01, *n* = 10.

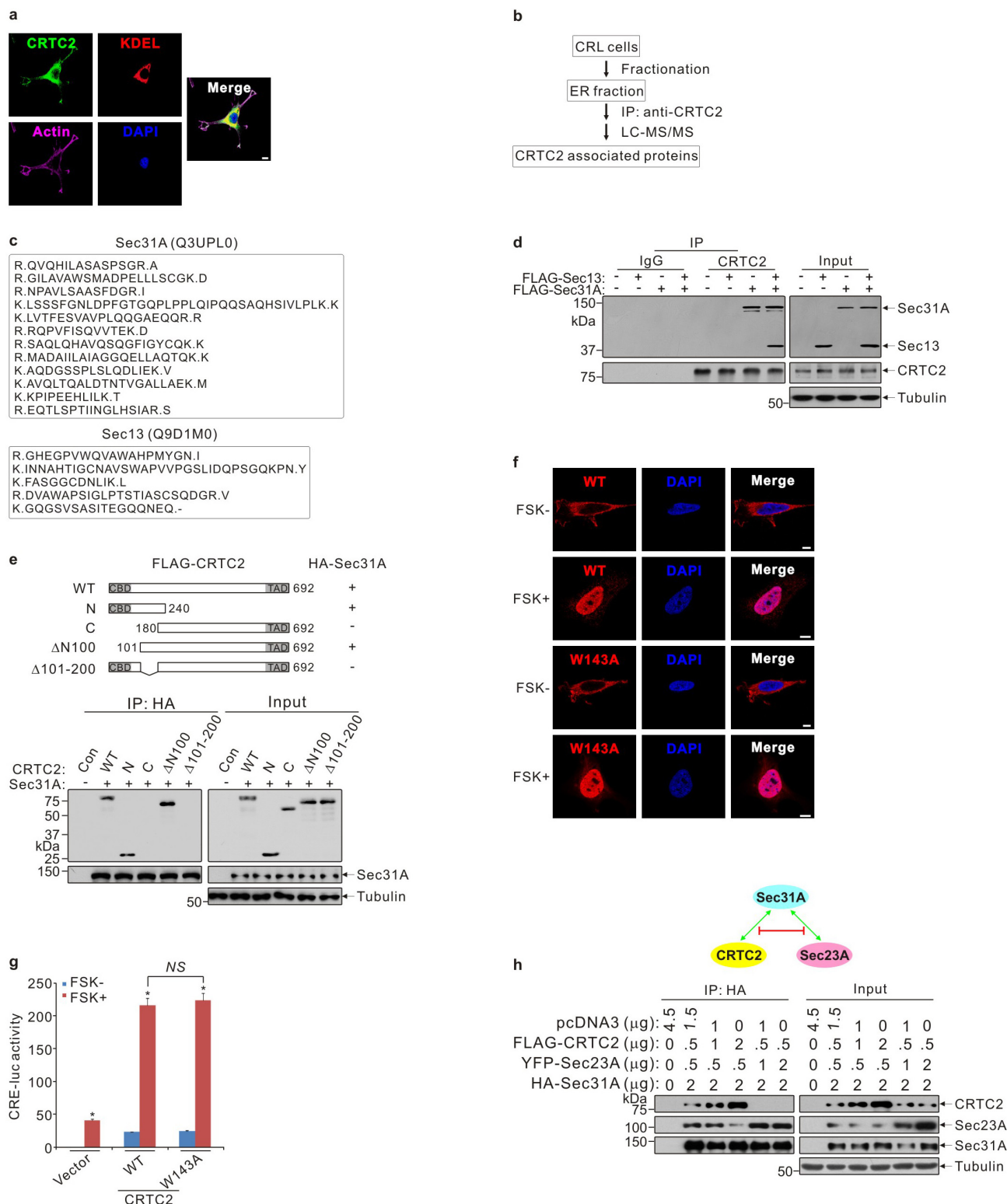


Extended Data Figure 2 | Validation of *Srebp1* knockdown in mice. **a–c**, Hepatic triglycerides (TGs) (**a**), immunoblots (**b**) and qPCR results for lipogenic gene expression (**c**) showing the effect of *Srebp1* RNAi in liver extracts of fed mice. Data are shown as mean \pm s.e.m. * $P < 0.01$, $n = 10$. US, unspecific.



Extended Data Figure 3 | Effect of CRTC2 and its mutants on gluconeogenic and lipogenic gene expression. **a**, Cellular localization of CRTC2 and its mutants CRTC2(ΔTAD) (amino acids 1–630) and CRTC2(ΔTAD/AA) (amino acids 1–630 with double alanine mutations at Ser171 and Ser275) in mouse primary hepatocytes. FSK, forskolin. Scale bar, 10 μm. **b**, Effect of CRTC2 and its mutants on Cre-luc activity in mouse

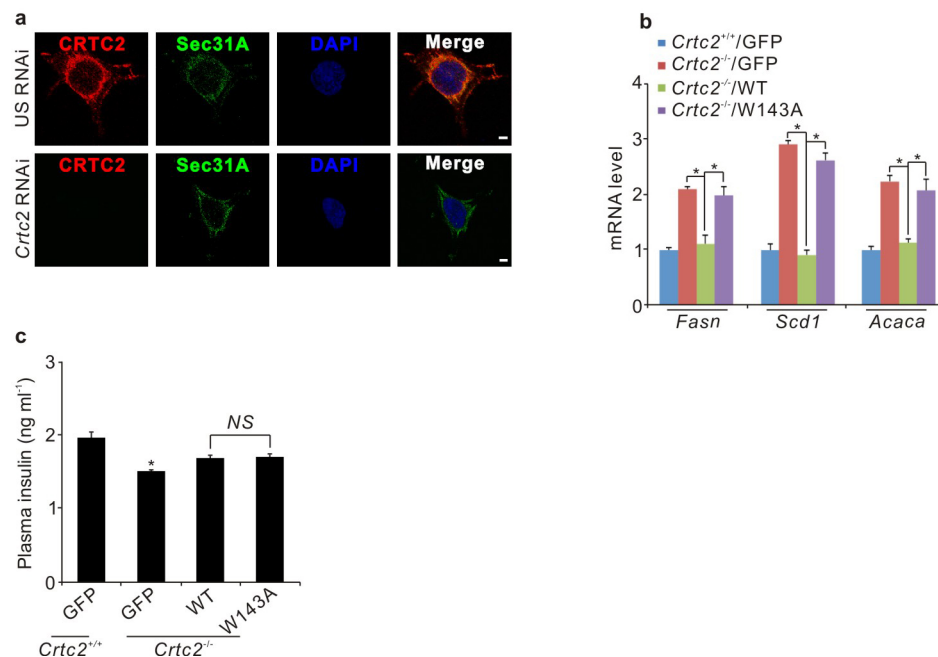
primary hepatocytes treated with or without 10 μM FSK for 4 h. Data are shown as mean ± s.e.m. * $P < 0.01$, $n = 6$. **c**, **d**, Effect of CRTC2 and its mutants on lipogenic gene (*Fasn*, *Scd1*, *Acaca*) expression (**c**) and plasma insulin level (**d**) in fed mice. Data are shown as mean ± s.e.m. * $P < 0.01$, $n = 8$. NS, no significant statistical difference.



Extended Data Figure 4 | Association of CRTC2 with Sec31A. **a**, Immunofluorescence showing relative co-localization of CRTC2 with an endoplasmic reticulum (ER) marker (KDEL) in CRL-2189 cells. Scale bar, 10 μm.

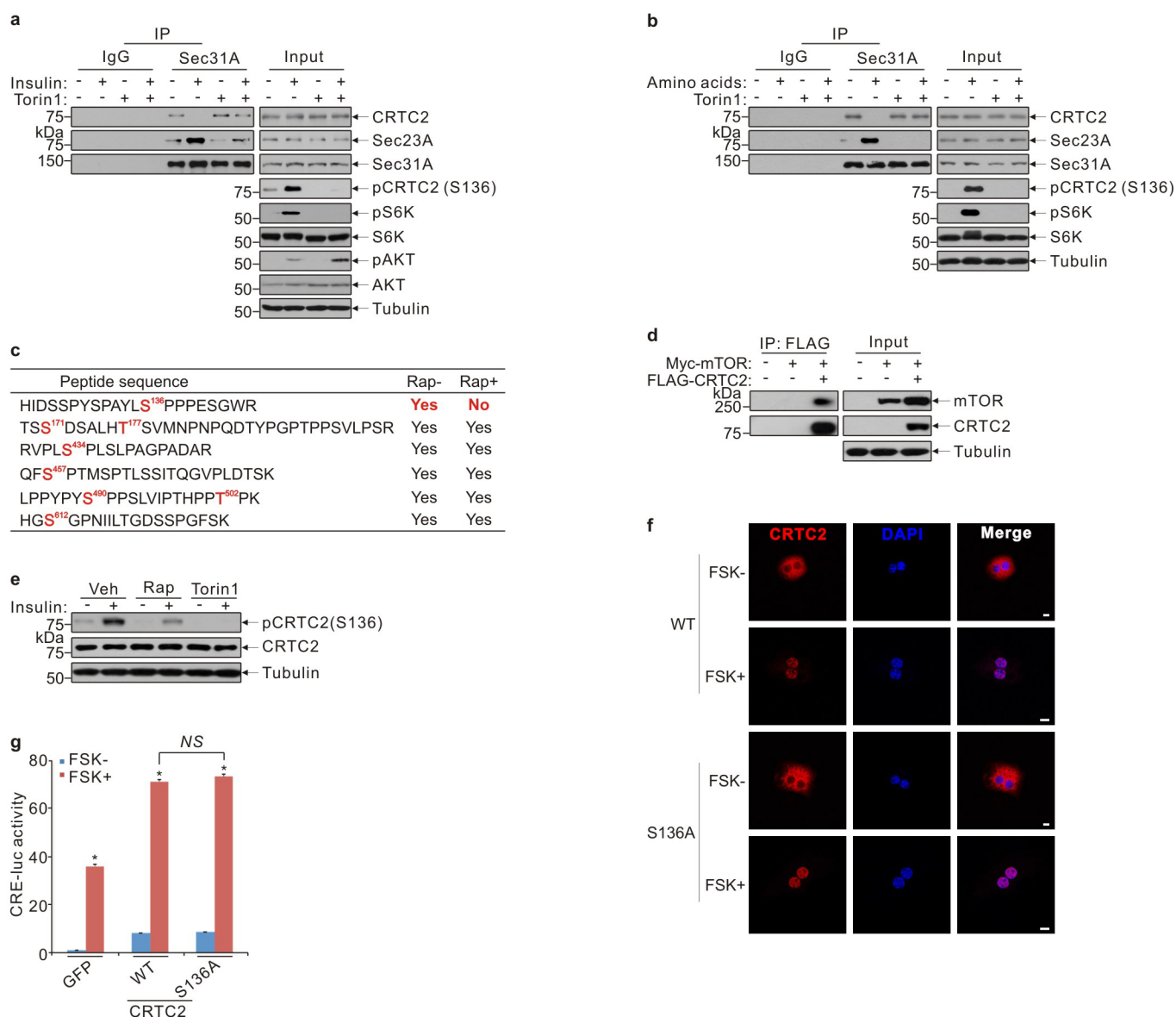
b, c, Strategy to purify CRTC2-interacting proteins (**b**), and the peptides identified from Sec31A and Sec13 (**c**) by MS analysis of immunoprecipitates prepared with anti-CRTC2 antibody from CRL-2189 ER fractions. **d**, Co-immunoprecipitation assay showing amounts of Flag-tagged Sec13 or Sec31A recovered from immunoprecipitates of endogenous CRTC2 in HEK293T cells.

e, Deletion analysis of regions in CRTC2 required for the CRTC2-Sec31A interaction. Interaction-competent CRTC2 peptides are indicated by (+) in each schematic. **f, g**, Cellular localization of the tryptophan-to-alanine mutant of CRTC2 (W143A) and its effect on Cre-luc activity in HEK293T cells. Scale bars, 10 μm. Data are shown as mean ± s.e.m. **P* < 0.01, *n* = 6. NS, no significant statistical difference. **h**, Co-immunoprecipitation assay showing amounts of Flag-tagged CRTC2 and YFP-tagged Sec23A recovered from immunoprecipitates of HA-tagged Sec31A in HEK293T cells.



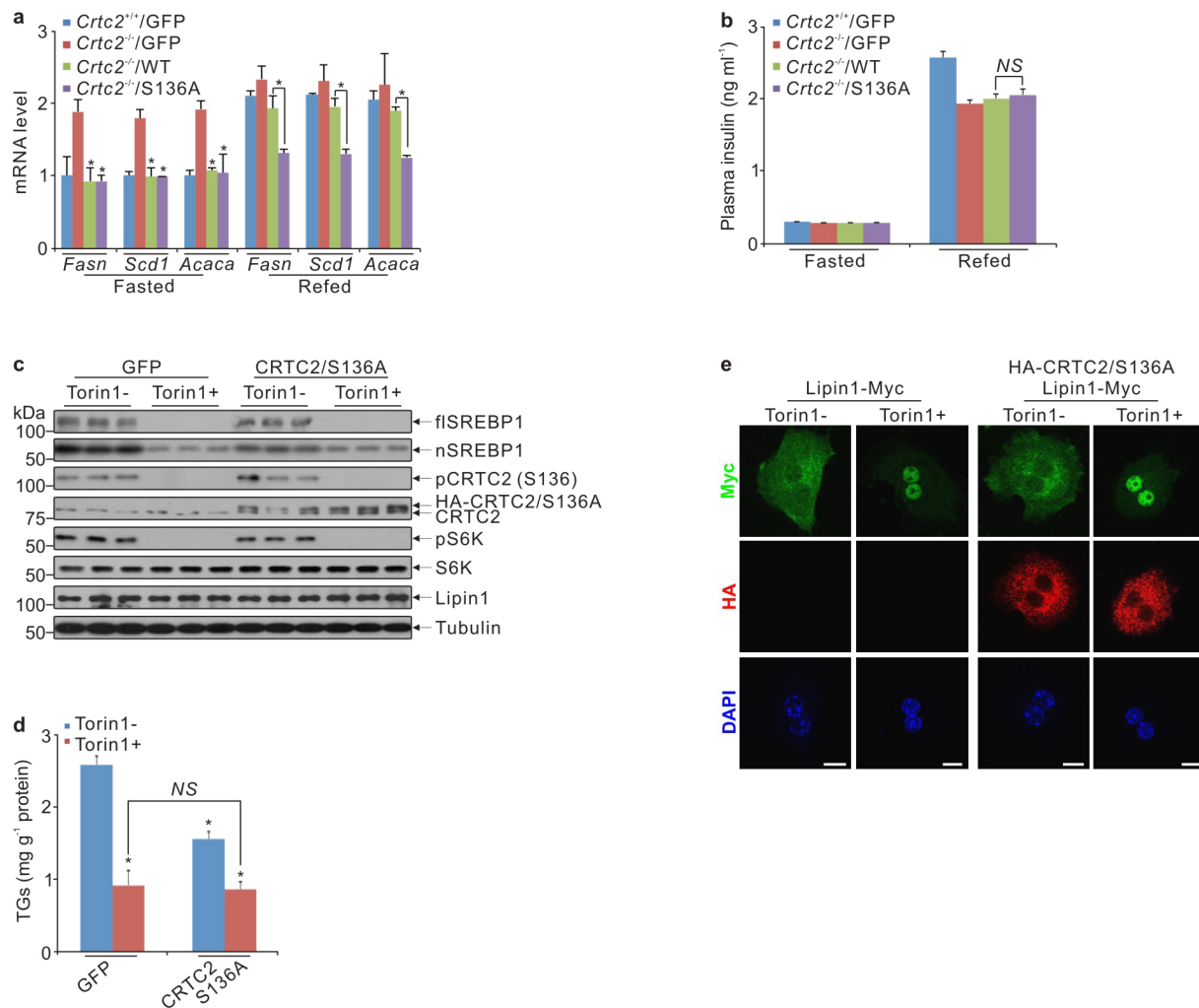
Extended Data Figure 5 | Modulation of COPII-dependent SREBP1 activity by CRTC2. **a**, Immunostaining showing the effect of *Crtc2* RNAi on the cellular localization of Sec31A. Scale bars, 10 μ m. **b**, **c**, Effect of wild-type (WT)

and Sec31A-interaction-defective (W143A) CRTC2 on lipogenic gene expression (**b**) and plasma insulin level (**c**) in fed mice. Data are shown as mean \pm s.e.m. * $P < 0.01$, $n = 8$. NS, no significant statistical difference.



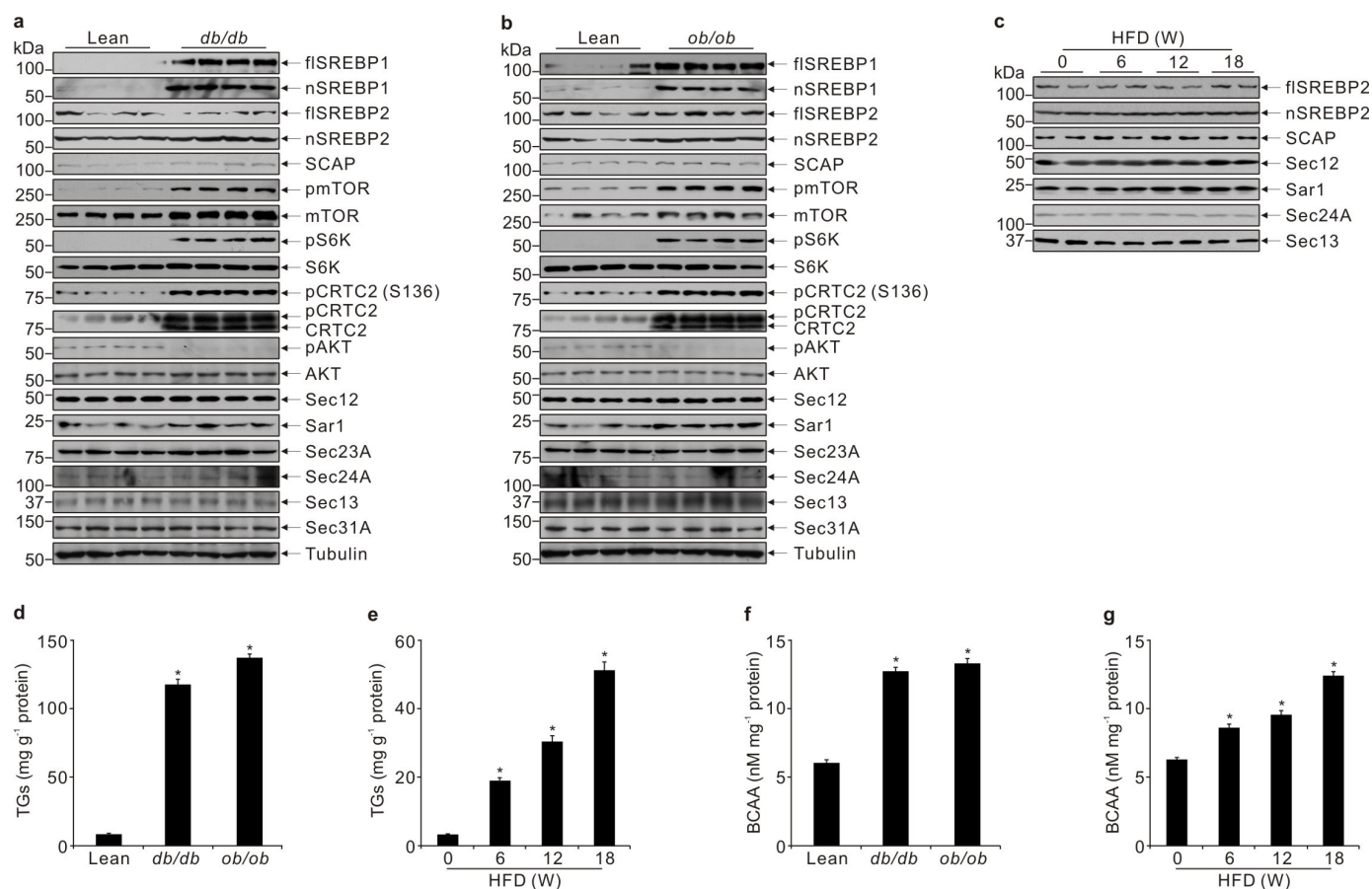
Extended Data Figure 6 | Characterization of CRTC2 phosphorylation site(s) by mTOR. **a**, Immunoblots showing co-immunoprecipitation of CRTC2 and Sec23A with Sec31A in mouse primary hepatocytes in response to insulin and/or torin1 treatment. Mouse primary hepatocytes were incubated with 250 nM torin1 or control vehicle for 1 h before 30 min insulin (100 nM) stimulation. Phospho-S6K (pS6K), total S6K, phospho-AKT (pAKT), total AKT and phospho-CRTC2 (Ser136) levels are also indicated. **b**, Immunoblots showing co-immunoprecipitation of CRTC2 and Sec23A with Sec31A in mouse primary hepatocytes in response to amino acids and/or torin1 treatment. Mouse primary hepatocytes incubated with amino-acid-free MEM for 3 h were exposed to 250 nM torin1 or control vehicle for another 1 h, then treated with amino acids for 30 min. **c**, Phospho-peptides of Flag-tagged CRTC2 identified by MS analysis of immunoprecipitates prepared with

anti-Flag from HEK293T cells treated with 100 nM rapamycin for 1 h (Rap+) or not (Rap-). Serine 136 was phosphorylated (Yes) in the absence of Rap treatment (Rap-) and dephosphorylated (No) in the presence of Rap (Rap+). **d**, Co-immunoprecipitation assay showing the association between Flag-tagged CRTC2 and Myc-tagged mTOR in HEK293T cells. **e**, Effect of the mTOR inhibitors Rap and torin1 on CRTC2 phosphorylation. Mouse primary hepatocytes were pretreated with vehicle (Veh), 100 nM Rap, or 250 nM torin1 for 1 h before 100 nM insulin stimulation for 30 min. **f**, **g**, Cellular localization of the phosphorylation-defective CRTC2 mutant (S136A) (**f**) and its effect on Cre-luc activity (**g**) in mouse primary hepatocytes. Scale bars, 10 μ m. Data are shown as mean \pm s.e.m. * P < 0.01, n = 6. NS, no significant statistical difference.



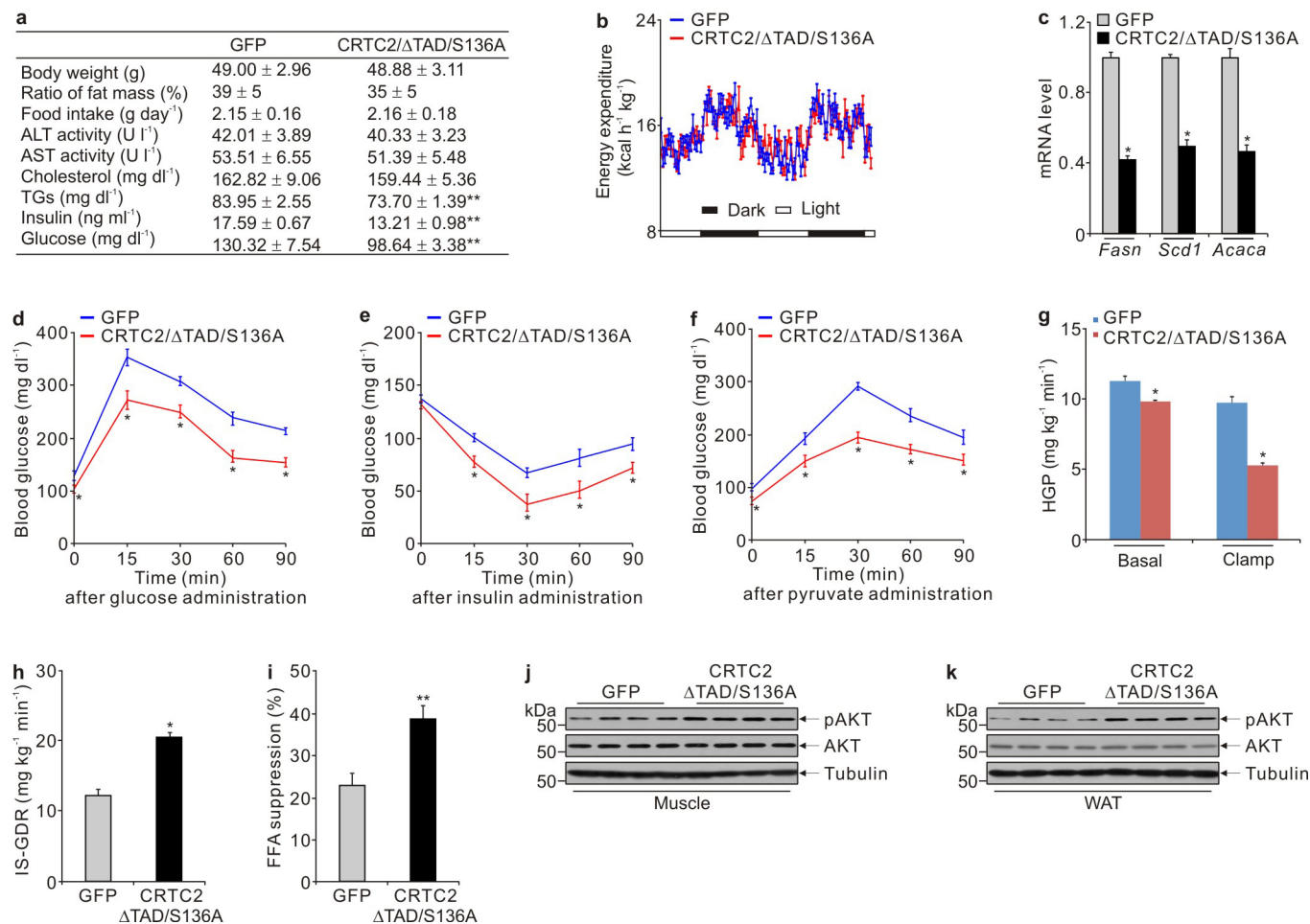
Extended Data Figure 7 | Effect of CRTC2(S136A) on SREBP1 maturation, lipin1 localization and circulating insulin level. **a, b**, Effect of wild-type or CRTC2(S136A) on lipogenic gene expression in liver (**a**) and plasma insulin level (**b**) of fasted (3 h) and refed (1 h after 3 h fasting) mice. **c–e**, Effect of CRTC2(S136A) and torin1 treatment on SREBP1 maturation (**c**), hepatic triglycerides (**d**) and lipin1 localization in mouse primary hepatocytes (**e**).

Torin1 (20 mg kg⁻¹) was intraperitoneally injected 6 h before livers were harvested. For lipin1 localization, mouse primary hepatocytes were treated with vehicle (Torin1–) or 250 nM torin1 (Torin1+) for 4 h. Scale bars, 10 μ m. Data are shown as mean \pm s.e.m. * P < 0.01, n = 8. NS, no significant statistical difference.



Extended Data Figure 8 | Enhanced SREBP1 activation, triglyceride levels and branched-chain amino acid levels in obese mice. **a–c**, Immunoblots showing relative amounts and/or phosphorylation status of SREBP1, SREBP2, SCAP, mTOR, S6K, CRTC2, AKT and COPII subunits in fed lean and *db/db* mice (**a**), *ob/ob* mice (**b**), and relative amounts of SREBP2, SCAP and COPII

subunits in HFD-fed mice (**c**). **d–g**, Hepatic triglyceride amounts and branched-chain amino acid (BCAA) levels in liver extracts from lean, *db/db*, *ob/ob* and HFD-fed mice in the fed state. Data are shown as mean \pm s.e.m. * P < 0.01, n = 10.



Extended Data Figure 9 | Improved insulin sensitivity in HFD-fed mice in the presence of CRTC2(Δ TAD/S136A). **a–k**, Effect of the mTOR-defective mutant CRTC2(Δ TAD/S136A) on metabolic parameters (**a**), including body weight, relative fat mass, food intake, plasma alanine aminotransferase (ALT) and aspartate aminotransferase (AST) activity, plasma cholesterol, plasma triglycerides, plasma insulin and blood glucose; energy expenditure (**b**);

lipogenic gene expression (**c**); glucose tolerance (**d**); insulin tolerance (**e**); pyruvate tolerance (**f**); hepatic glucose production (HGP; **g**); insulin-stimulated glucose disposal rate (IS-DGR; **h**); percentage of free fatty acid (FFA) suppression (**i**); pAKT level in skeletal muscle (**j**); and pAKT level in epididymal white adipose tissue (**k**) from mice fed on a HFD for 18 weeks. Data are shown as mean \pm s.e.m. * P < 0.01, ** P < 0.05, n = 8 (**a–f**), n = 6 (**g–i**).

Crucial HSP70 co-chaperone complex unlocks metazoan protein disaggregation

Nadinath B. Nillegoda¹, Janine Kirstein², Anna Szlachcic¹, Mykhaylo Berynskyy³, Antonia Stank^{3,4}, Florian Stengel⁵, Kristin Arnsburg², Xuechao Gao¹, Annika Scior², Ruedi Aebersold^{5,6}, D. Lys Guilbride¹, Rebecca C. Wade^{1,3,7}, Richard I. Morimoto⁸, Matthias P. Mayer¹ & Bernd Bukau¹

Protein aggregates are the hallmark of stressed and ageing cells, and characterize several pathophysiological states^{1,2}. Healthy metazoan cells effectively eliminate intracellular protein aggregates^{3,4}, indicating that efficient disaggregation and/or degradation mechanisms exist. However, metazoans lack the key heat-shock protein disaggregase HSP100 of non-metazoan HSP70-dependent protein disaggregation systems^{5,6}, and the human HSP70 system alone, even with the crucial HSP110 nucleotide exchange factor, has poor disaggregation activity *in vitro*^{4,7}. This unresolved conundrum is central to protein quality control biology. Here we show that synergic cooperation between complexed J-protein co-chaperones of classes A and B unleashes highly efficient protein disaggregation activity in human and nematode HSP70 systems. Metazoan mixed-class J-protein complexes are transient, involve complementary charged regions conserved in the J-domains and carboxy-terminal domains of each J-protein class, and are flexible with respect to subunit composition. Complex formation allows J-proteins to initiate transient higher order chaperone structures involving HSP70 and interacting nucleotide exchange factors. A network of cooperative class A and B J-protein interactions therefore provides the metazoan HSP70 machinery with powerful, flexible, and finely regulatable disaggregase activity and a further level of regulation crucial for cellular protein quality control.

To investigate the possibility of a potent protein disaggregation activity in metazoans, we focused on the HSP70 chaperone system, which displays some *in vitro* capacity to disentangle and refold aggregated polypeptides when powered by an HSP110 co-chaperone^{4,7}. The HSP70–J-protein–HSP110 functional cycle described in Extended Data Fig. 1a, by generally accepted extrapolation, occurs on protein aggregate surfaces. Homodimeric J-proteins are essential components of this cycle^{8,9}. Three classes of J-proteins (A, B and C) with >50 members in humans, determine HSP70 substrate selection, with some functional redundancy among members⁹. For example, class A and B J-proteins (Fig. 1a) implicated in protein quality control have common functions, but independent and differing efficacies^{9–11}. The basis for the evolutionary maintenance of these two classes of J-proteins (despite appreciable internal diversity^{12,13}), and the relation of class to function and principles governing substrate selection, remain unknown.

Here we explore the full potential of the metazoan HSP70–J-protein–HSP110 system in protein disaggregation, by examining the functional relationship between class A and B J-proteins. Using thermally denatured luciferase from *Photinus pyralis* as model substrate⁴, we investigate the *in vitro* protein disaggregation/refolding versus protein refolding-only (Extended Data Fig. 1b–d) capacities of the human and *Caenorhabditis elegans* HSP70–HSP110 systems (also known as HSPA8–HSPH2 in humans, and HSP-1–HSP-110 in *C. elegans*) in

conjunction with class A and B J-proteins (Fig. 1, Extended Data Fig. 1e and Extended Data Table 1).

In disaggregation/refolding reactions with a class A (JA2) and B (JB1) J-protein present together (Fig. 1b, magenta), rather than either class J-protein alone (Fig. 1b, green or blue), we observed unprecedented reactivation of pre-formed heat-aggregated luciferase, indicating synergistically accelerated protein disaggregation. This was also seen under limiting chaperone concentrations (maintained in all further experiments) with multiple class A (JA1 and JA2) and class B (JB1 and JB4) human J-proteins (Extended Data Figs 1f and 2a). Disaggregation reactions with the corresponding nematode HSP-1–HSP-110 system and J-proteins DNJ-12 (class A) and DNJ-13 (class B) show similar synergic acceleration (Fig. 1c and Extended Data Fig. 2c, d). In reactions containing only one J-protein class (Extended Data Fig. 1f, JA2, solid lines; or JB1, dashed lines), with increased J-protein levels of threefold or more relative to the mixed-class J-protein reaction (Extended Data Fig. 1f, magenta), protein disaggregation/refolding slows and is inhibited. We infer that the presence of class A and B J-proteins together, rather than J-protein amount, determines reaction efficiency. Both the disaggregation/refolding rate (Extended Data Fig. 2e, f) and yield (Extended Data Fig. 2g) of renatured luciferase peak with equal proportions of class A to B. A broad range of flanking reciprocal A to B J-protein stoichiometries also show appreciable activity, suggesting that efficient disaggregation/refolding is supported by minimal amounts of preferentially interacting A and B J-proteins. Increased initial rates at higher stoichiometries of JA2 (Extended Data Fig. 2f) reflect intrinsically higher refolding capacity of class A J-proteins with HSP70 (Extended Data Fig. 2b, green)¹⁴.

Disaggregation synergy in mixed J-protein class reactions occurs with and without small HSP (*Saccharomyces cerevisiae* Hsp26) incorporation into aggregates for both human (Extended Data Figs 1f and 3a) and nematode J-protein containing systems (Fig. 1c and Extended Data Fig. 2d). Synergy is independent of nucleotide exchange factors (NEFs) (Extended Data Fig. 3b), protein substrate (Fig. 1b and Extended Data Fig. 3c, d) and substrate concentration variations affecting density, size⁴ and therefore the architectural nature of the aggregate generated (Extended Data Fig. 3e). Synergy also occurs at lower chaperone to substrate ratios (Fig. 1b and Extended Data Figs 1f and 3f), and at different and characteristic ranges of substrate to J-protein ratio for malate dehydrogenase (MDH) versus luciferase or α -glucosidase disaggregation (Fig. 1b and Extended Data Fig. 3c, d). MDH aggregates resolve considerably with non-limiting concentrations of JB1 alone (not shown), but with limiting JB1 concentrations in the presence of JA2, synergic MDH disaggregation occurs (Extended Data Fig. 3d). Synergy in disaggregation therefore appears generic, operating over a range of ratios and concentrations, with room for substrate-linked variation. By contrast, refolding-only reactions show

¹Center for Molecular Biology of the University of Heidelberg (ZMBH), German Cancer Research Center (DKFZ), DKFZ-ZMBH Alliance, 69120 Heidelberg, Germany. ²Leibniz-Institute for Molecular Pharmacology (FMP), 13125 Berlin, Germany. ³Heidelberg Institute for Theoretical Studies (HITS), 69118 Heidelberg, Germany. ⁴Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences, Heidelberg University, 69120 Heidelberg, Germany. ⁵Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, 8093 Zurich, Switzerland. ⁶Faculty of Science, University of Zurich, 8057 Zurich, Switzerland. ⁷Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, 69120 Heidelberg, Germany. ⁸Department of Molecular Biosciences, Rice Institute for Biomedical Research, Northwestern University, Evanston, Illinois 60208, USA.

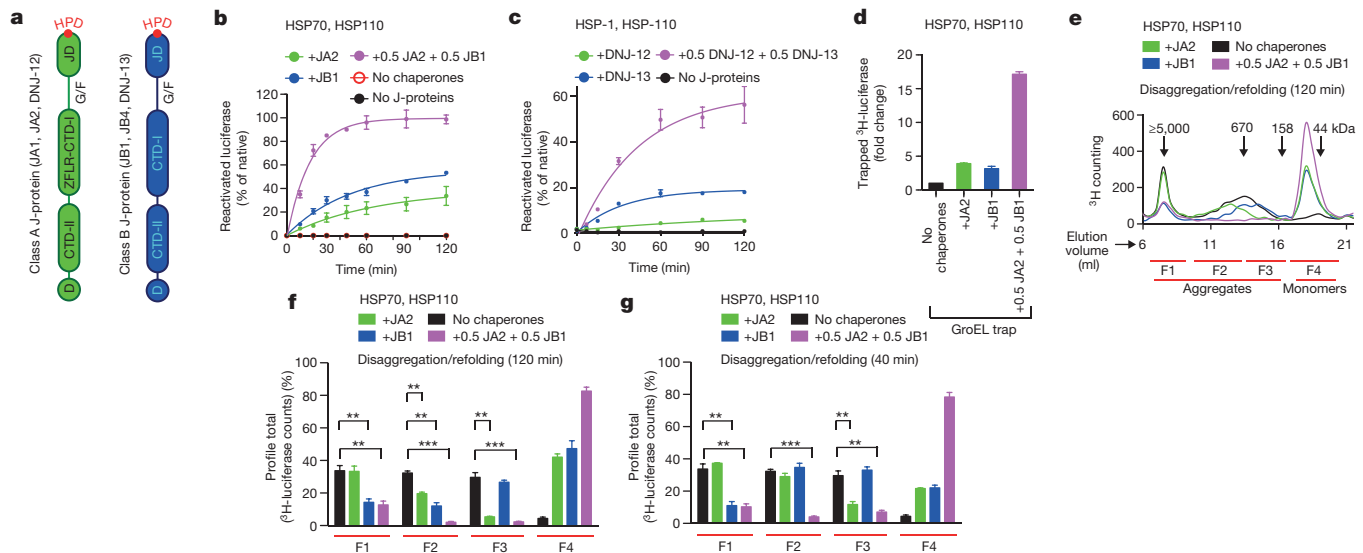


Figure 1 | Simultaneous presence of class A and B J-proteins unleashes protein disaggregation activity and broadens target aggregate range of the HSP70 machinery. **a**, Two distinct classes (A and B) display highly conserved domain organization involving the HSP70-interacting HPD motif (red) containing amino-terminal J-domain (JD), Gly/Phe-rich flexible region (G/F), C-terminal β-sandwich domains (CTD-I and II), with class A J-proteins distinguished mainly by a zinc-finger-like region (ZFLR) that inserts into the CTD-I subdomain and a dimerization domain (D)^{9,23}. CTD together with ZFLR provide substrate specificity^{24,25}. **b**, Disaggregation and reactivation of preformed luciferase aggregates using human HSP70–HSP110 with human J-proteins JA2 (green), JB1 (blue), JA2+JB1 (magenta) or with no J-proteins (black) (n = 3). **c**, Reactivation of heat-aggregated luciferase by nematode HSP70 machinery

containing HSP-1, HSP-110 and either alone or in combination with the nematode J-proteins DNJ-12 (A) and DNJ-13 (B) (n = 2). **d**, Fold change in trapped luciferase; control, GroEL^{D87K} without other chaperones (black). Values normalized to total ³H counts in each reaction (n = 2). **e**, SEC profile after disaggregation/refolding (120 min) with either J-protein alone or combined. Elution fractions labelled F1–F4 (red lines); F4, disaggregated monomers (~63 kDa). **f**, Aggregate quantification from fractions F1–F4 from the SEC profile in **e**. Disappearance of ³H-luciferase from aggregates (F1–F3) occurs with concomitant accumulation of disaggregated monomer (F4). **g**, Aggregate quantification, after 40-min disaggregation. Values normalized to total counts in each reaction. Two-tailed t-test, **P < 0.01, ***P < 0.001 (n = 3). Data are mean ± s.e.m. Precise concentrations are shown in Extended Data Table 1.

no synergism (Extended Data Fig. 2b). We conclude that efficient protein disaggregation, but not refolding, requires cooperation between class A and B J-proteins.

Three non-exclusive mechanisms could explain the synergistic action of class A and B J-proteins. In a mechanism involving sequential action, one J-protein class interacts with HSP70–HSP110 to extract polypeptides from aggregates. The other J-protein class then prevents re-aggregation of extracted polypeptide (holdase function) and/or in combination with HSP70–HSP110 promotes substrate refolding. Of the four J-proteins tested for holdase function, only JA2 and JB4 prevent luciferase aggregation at 42 °C (Extended Data Fig. 3g, h). However, disaggregation synergy is indistinguishable for J-protein combinations with (JA2 or JB4) and without (JA1 or JB1) holdase function (Extended Data Fig. 2a). Furthermore, disaggregation/refolding rates are unaffected by the order of JA2 and JB1 addition during the reaction (Extended Data Fig. 3i), indicating that J-proteins act in no strict order. For direct validation, we quantified tritium-labelled luciferase extracted from aggregates using a mutant GroEL protein (GroEL^{D87K}) as a trap¹⁵ for extracted luciferase molecules, preventing refolding. Decreased luciferase activity in disaggregation/refolding reactions in the presence of GroEL^{D87K} reflects trapping of labelled disaggregated polypeptides (Extended Data Fig. 4a, b), counted by measuring tritium scintillation (Fig. 1d). Disaggregation/refolding reactions containing only one class of J-protein show similar amounts of trapped ³H-labelled luciferase polypeptides. With class A and B J-proteins present together, however, we see synergistically accelerated accumulation of disaggregated ³H-labelled luciferase trapped in GroEL (Fig. 1d). Together, these results exclude a strictly sequential function of J-protein classes in disaggregation/refolding, corroborating the inference that synergy occurs at the protein disaggregation step.

A second model stipulates that each J-protein class acts specifically, in parallel, distinguishing protein aggregates by size and/or compactness during the disaggregation step. We tested this by adding different J-protein–HSP70–HSP110 mixtures to preformed ³H-labelled luciferase aggregates, which display a range of sizes, and probably variations

in molecular architecture. We analysed the disaggregation of aggregate populations by size-exclusion chromatography (SEC; Fig. 1e–g and Extended Data Fig. 4c, d). Reactions were run in parallel, stopped by depleting ATP with apyrase, and held on ice until SEC (Extended Data Fig. 4e). Eluted fractions (F1–F4, Fig. 1e–g) reveal JA2-containing chaperone mixes preferentially solubilize smaller aggregates (F3; ~200–700 kilodaltons (kDa)). Conversely, JB1-containing mixes preferentially solubilize larger aggregates (F1, ≥5,000 kDa; F2, ~700–4,000 kDa), but solubilize small aggregates less efficiently. These results are consistent with distinct, parallel class activity. JA2 plus JB1 combinations, however, in much shorter reactions (40 min instead of 120 min), solubilize both larger and smaller aggregates far more efficiently than the added efficiencies of separate JA2 and JB1 reactions allow (Fig. 1g). Similar results obtain throughout for α-glucosidase aggregate solubilization (Extended Data Fig. 4d). This suggests concerted action on the same target.

This prompts a third model, in which synergy results from the formation of mixed-class J-protein complexes exerting concerted activity to facilitate disaggregation. A range of approaches rigorously tests this model.

To visualize individual versus complexed J-protein function, we biased disaggregation/refolding reactions by combining JA2:JB1 in 5:1 to 1:5 ratios, then analysed aggregate resolution by SEC (Extended Data Fig. 4f). The 1:1 ratios dissolve all aggregates (F1–F3, magenta). In contrast, limiting JB1 concentration and excess JA2 in shorter reactions (40 min, orange solid) barely resolves the largest aggregates (F1), whereas the smaller aggregates (F2–F3) disappear completely; F1 aggregates resolve only in longer reactions (120 min, orange hash). Limiting JB1 concentrations alone, however, readily resolve large F1 aggregates (blue solid). We infer that scarce JB1 molecules preferentially sequester with excess JA2 into complexes that efficiently process all sizes of aggregates; the smaller F2 and F3 aggregates accordingly disappear first. Reciprocal titration with scarce JA2 and excess JB1 concentration shows less disaggregation of the smaller F2 and F3 aggregates (magenta versus red

solid, Extended Data Fig. 4f), which fully resolve with a longer reaction time (120 min, red hash). Specific J-protein stoichiometries evidently modulate HSP70 targeting and disaggregation efficacy. We infer that J-proteins preferentially form efficient mixed-class complexes, supporting a model for concerted action.

Independent tests for physical interactions between class A and B J-proteins consistently reveal intermolecular J-domain–C-terminal-domain (JD–CTD) and CTD–CTD contacts. Approaches include chemical cross-linking coupled to mass spectrometry (Fig. 2a), Förster resonance energy transfer (FRET; Fig. 2b), docking simulations (Fig. 2c, d) and competition assays (Fig. 2e).

Mass spectrometry of JA2 and JB1 combinations treated with lysine-specific cross-linker (disuccinimidyl suberate) identifies three intermolecular cross-linked regions between JD^{JA2}–CTD^{JB1}, JD^{JB1}–CTD^{JA2} and CTD^{JA2}–CTD^{JB1} (Fig. 2a and Extended Data Fig. 5a, b). FRET measured by donor quenching indicates JD–CTD and CTD–CTD interactions between JA2 and JB1 in solution (Fig. 2b, J-protein pairs 1, 2 and 3; Extended Data Fig. 6a). This corroborates our cross-linking data and favours biological relevance. We detect neither JD–JD interactions between classes (J-protein pair 4), nor intermolecular same-class JD–CTD interactions (J-protein pair 5). However, in agreement with structures from small-angle X-ray scattering of class B J-proteins^{16,17}, we detect JD^{JB1}–CTD^{JB1} cross-links (not shown). Presumably these reflect intramolecular interactions, preventing intermolecular JD^{JB1}–CTD^{JB1} but not JD^{JA2}–CTD^{JB1} interactions, as indicated by FRET (Fig. 2b).

We further defined the interface of the JA2–JB1 complex using unbiased docking simulations between J-domain and CTD dimers of JA1, JA2, JB1 and JB4 (Fig. 2c, d and Extended Data Fig. 7a, b). Simulations show a preferred binding arrangement of JD^{JB1} on CTD^{JA2} and conversely JD^{JA2} on CTD^{JB1} (Fig. 2c, d), again corroborating cross-linking data (Fig. 2a).

Furthermore, in competition experiments, the addition of moderate excess of isolated J-domain fragments inhibits JA2–JB1–HSP70–HSP110-dependent disaggregation/refolding of heat-aggregated luciferase (Fig. 2e), although not refolding alone (Fig. 2f). J-domain fragments carrying the HPD motif mutated to QPN, which abolishes the JD–HSP70 interaction and ATP hydrolysis stimulation on HSP70 (refs 18, 19), have the same effect (Extended Data Fig. 6e), confirming that inhibition of disaggregation is not due to HSP70 being titrated out by J-domain fragment binding. Unlabelled full-length J-proteins and isolated J-domains compete with mixed-class JD–CTD interactions, indicated by decreased FRET efficiency between JA2 and JB1

(Extended Data Fig. 6f, g), explaining the inhibitory effects. However, JD–CTD interaction sites do not overlap CTD binding sites for substrate, since JA2 holdase activity remains unaffected with isolated J-domains present (Extended Data Fig. 7c, d). Molecular docking modelling supports this also (Extended Data Fig. 8). J-protein complexing involving mixed-class J-domains and CTDs is therefore crucial for efficient protein disaggregation, but not for refolding.

Non-ionic detergent affects neither disaggregation activity (Extended Data Fig. 6b) nor FRET efficiency between class A and B molecules (Extended Data Fig. 6a). Increasing salt concentrations, however, weaken both (Extended Data Fig. 6c, d), suggesting ionic interactions. Independent methodologies therefore confirm specific JD–CTD interactions of a predominantly electrostatic nature directly implicated in disaggregation efficiency.

J-domain and CTD regions display highly conserved, class-specific electrostatic potentials (Fig. 3a, b). Class A J-proteins show distinct polarity in the CTDs, with negatively charged regions (red) in the CTD-II and dimerization subdomains, and positively charged regions (cyan) along the zinc-finger-like region and CTD-I hook (Fig. 3a). Conversely, class B CTDs are relatively non-polar, with positively charged regions in the CTD where JD^{JA2} cross-linking occurs. J-domains in both classes are markedly bipolar, although class A J-domains have smaller negatively charged regions (Fig. 3b). In all J-domains, positive charge (near the HPD motif and helix-II) is implicated in binding to HSP70 (refs 18, 19). We deduce conserved negatively charged regions exposed in the J-domains interact with positively charged CTD regions in opposite class J-proteins.

We therefore generated triple charge-reversal variants of the J-domain (JA2^{RRR} or JB1^{RRR}), replacing negatively charged Asp or Glu residues with positively charged Arg residues in and around helices-I and -IV (Fig. 3c). FRET interactions between the JD^{JA2} and CTD^{JB1} regions diminish with charge-reversal mutations in either JD^{JA2} or JD^{JB1} (Fig. 3d, J-protein pairs 2 and 3), and are abrogated with charge reversals in both interacting J-domains (Fig. 3d, J-protein pair 4). Partial FRET reduction with triple charge reversals in only one interacting JD–CTD domain pair suggests some degree of intermolecular tethering by the other pair, although insufficient for full J-protein cooperation and disaggregation efficiency (Fig. 3e). In refolding-only reactions, recovered luciferase activity remains unaffected by J-domain charge reversals (Fig. 3f). Physically complexed and cooperating mixed-class J-proteins are therefore essential for efficient HSP70-dependent disaggregase activity, and are thought (but not directly shown) to act on the surface of

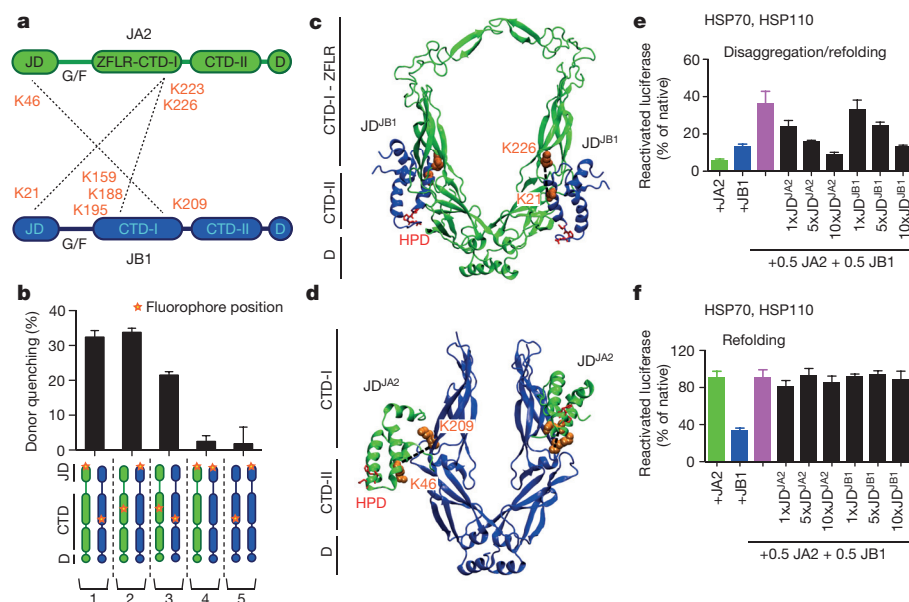


Figure 2 | Intermolecular JD–CTD interaction is required for mixed-class J-protein complex formation. **a**, Intermolecular cross-links (dashed lines) between Lys residues (orange) on JA2 (green) and JB1 (blue). **b**, JA2 and JB1 interactions analysed by FRET. Bars show donor quenching efficiency of JD–CTD interactions; cartoons below show fluorophore positions in J-protein protomer pairs 1–5. N-termini of JD^{JA2} and JD^{JB1} are labelled with acceptor fluorophore ReAsH. CTD^{JA2} and CTD^{JB1} are labelled with donor fluorophores FLAsH and Alexa Fluor 488 at residues 241 and 278, respectively ($n = 3$). **c**, **d**, Ribbon diagrams showing representative positions of JDs on CTD dimers from docking simulations; cross-linked Lys residues (space filling, orange, connected with black dashed lines) established in **a**; HPD motif (stick representation, red). **c**, JD^{JB1} (blue) and CTD^{JA2} (green). **d**, JD^{JA2} (green) and CTD^{JB1} (blue). **e**, **f**, Competition of excess isolated JD fragments for classes A and B J-protein complex formation and effect on luciferase disaggregation. **e**, **f**, Protein disaggregation/refolding (**e**) and refolding-only (**f**) ($n = 3$). Data are mean \pm s.e.m. Precise concentrations are shown in Extended Data Table 1.

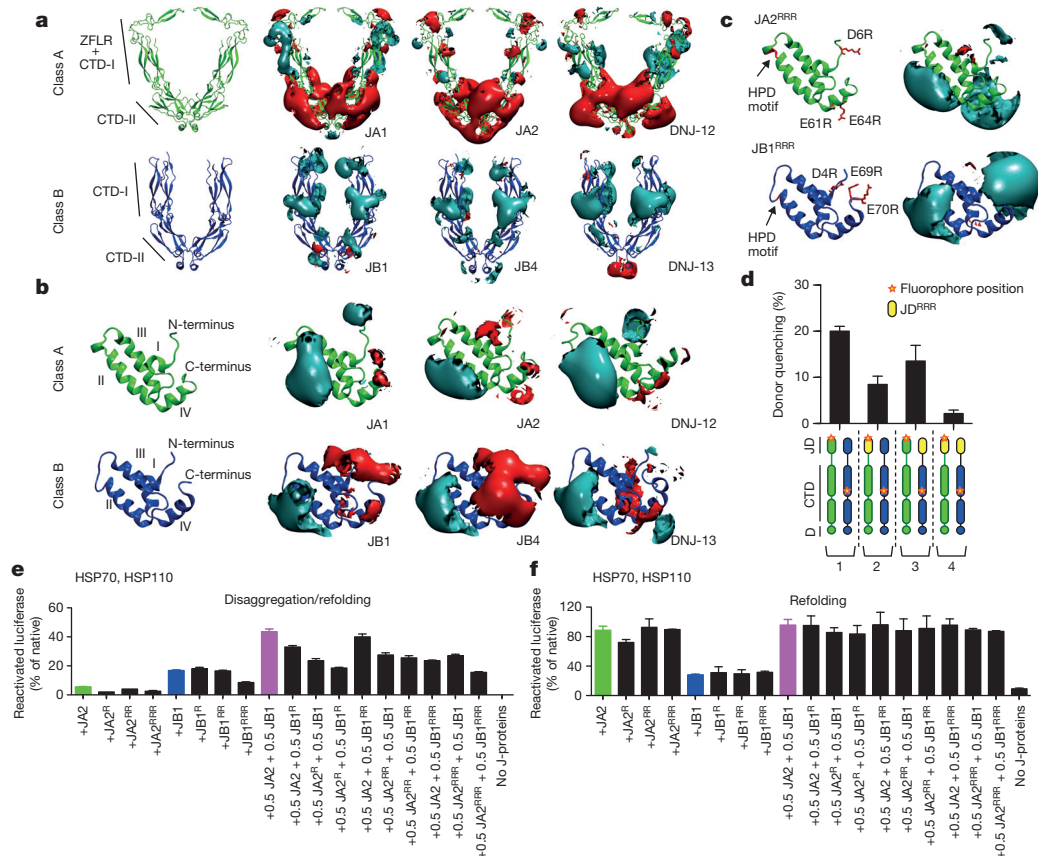


Figure 3 | Conserved electrostatic potential distributions in A and B J-protein classes are complementary and direct mixed-class J-protein interactions for complex formation. **a**, Electrostatic isopotential maps of CTD dimers comparing human (JA1, JA2, JB1 and JB4) and nematode (DNJ-12 and DNJ-13) class A and B J-proteins. Electrostatic potential around proteins is contoured at +1 (positive, cyan) and -1 (negative, red), kcal mol⁻¹ e⁻¹. Protein structures are represented by ribbon diagrams. **b**, Conserved α -helices and electrostatic isopotential maps contoured as in **a** of human and nematode J-domains. I–IV (from N-terminus) denote conserved α -helices. **c**, The J-domains of charge-reversal triple mutants (JA2^{RRR} and JB1^{RRR}); and their electrostatic

aggregates. A paradoxical sequence dispensability of these highly conserved helices-I and -IV observed in early data, which assayed exclusively for HSP70 interaction and protein folding^{18,20}, is also now explained. These data together strongly support a mixed-class J-protein interaction with vital function conserved in evolution.

Size separation of tritiated JB1 mixed with unlabelled, larger JA2, or the reciprocal labelling, reveals only JB1 (blue) or JA2 (green) homo-

isopotential maps compare with wild-types in **b**. RRR denotes triple amino acid substitutions D6R, E61R and E64R in JA2, and D4R, E69R and E70R in JB1. **d**, FRET determination of JA2 and JB1 triple charge-reversal mutants ($n = 3$). Bars show donor quenching efficiency of JD–CTD interactions; cartoons below show fluorophore positions in J-protein protomer pairs 1–4. Triple charge mutants are yellow. **e**, Luciferase disaggregation/refolding at 120 min with J-domain charge-reversal mutants (JA2: D6R (R); E61R + E64R (RR); D6R + E61R + E64R (RRR). JB1: D4R (R); E69R + E70R (RR); D4R + E69R + E70R (RRR)) ($n = 3$). **f**, As in **e**, refolding-only at 80 min ($n = 3$). Data are mean \pm s.e.m. Precise concentrations are shown in Extended Data Table 1.

dimers (Extended Data Fig. 5c), indicating that J-protein complexes are transient. Transient interactions would support an HSP70 disaggregation machinery with a flexible range of tailored activities. Single-class J-protein function shows HSP70–HSP110-mediated disaggregation activity limited to aggregates of specific size ranges (Fig. 4, large or small aggregates). Mixed-class J-protein complexes efficiently disaggregate a wide range of aggregate sizes (Fig. 4, large, medium and

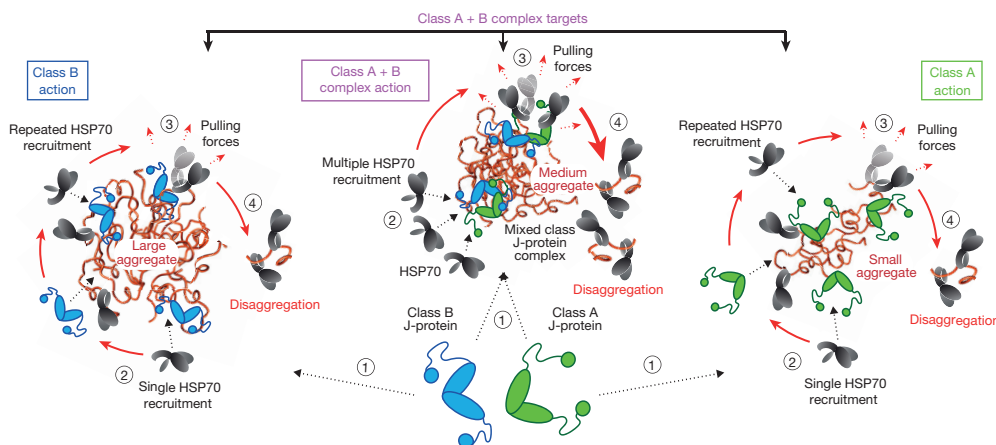


Figure 4 | Model of individual versus complexed class A and class B J-protein function in protein disaggregation. Size-specific aggregate targeting: large aggregates are targeted by J-protein^{class B}–HSP70–HSP110 (blue); small aggregates are targeted by J-protein^{class A}–HSP70–HSP110 (green); all aggregate sizes are targeted by J-protein-mixed-class-complex–HSP70–HSP110 (magenta). HSP70 molecules are in grey. Sequential reaction steps (encircled numbers): 1, J-protein targets aggregate; 2, J-protein recruits HSP70; 3, surface-bound chaperones generate pulling forces (dashed red arrows); and 4, polypeptide extraction leading to protein disaggregation. Chaperone recruitment denoted by dashed black arrows.

small). On the basis of our results, we reason a minimum complex consists of one class A J-protein homodimer binding to one class B homodimer in a 1:1 ratio, indicating that there are four J-domains per complex. Assuming two J-domains engage in interactions sufficient to complex the J-proteins, one J-domain per homodimer remains free to interact with one HSP70, allowing for recruitment of two interacting HSP70 molecules per complex without steric hindrance (Fig. 4, medium aggregates, Extended Data Fig. 8). We conclude that each mixed class J-protein complex recruits at least two HSP70 molecules per targeting event, possibly seeding dynamic, higher order chaperone assemblies on aggregate surfaces.

Our computational models of the structures of mixed class J-protein complexes (Extended Data Fig. 8) incorporate the constraints defined by all our cross-linking, FRET and docking data. In each model, space in the J-protein complex allows for substrate binding via several interfaces, HSP70 interaction with J-domains, and HSP110 interaction with each HSP70 protein. These models accommodate the concept of entropic pulling, in which HSP70 binding to entangled polypeptides decreases entropy, generating reciprocal forces that pull polypeptides from aggregates²¹. Such higher order chaperone complexes would be expected to increase pulling forces and stabilize disaggregating polypeptides by providing increased substrate binding surface, thereby accelerating protein disaggregation (Fig. 4; class A + B complex). Although also likely, direct verification of mixed-class J-protein–HSP70 complexes interspersed with single-class J-protein–HSP70 complexes on aggregate surfaces is currently experimentally intractable.

In summary, we demonstrate potent protein disaggregation activity in metazoans, mediated by the central HSP70–J-protein–HSP110 chaperone network. Disaggregation efficacy comparable to that of non-metazoan HSP100–HSP70 bi-chaperone systems, over a broad aggregate size range, requires transient physical interaction between class A and B J-proteins. The assembly of higher order chaperone complexes on protein aggregate surfaces is expected to increase coordinated pulling forces on multiple trapped polypeptides, providing a plausible mechanistic basis for increased disaggregation efficacy. Mixed-class J-protein complexes form preferentially and interact with HSP70–HSP110 to resolve a broad range of aggregates efficiently, whereas single-class J-protein–HSP70–HSP110 interaction targets specific aggregate sizes. This suggests intracellular J-protein stoichiometry will differentially regulate HSP70-dependent protein disaggregation efficiency. The transitory nature of J-protein complexes would, in this context, facilitate flexible response according to need. As in nematodes, human cytosol contains several members of J-protein classes: four class A and nine class B J-proteins⁹. A wide range of complexed J-protein combinations is therefore available in humans and other metazoa, providing flexible target selectivity. This opens the further possibility of physiological function in assembly/disassembly of other macromolecular cell structures. These findings may also impinge on the amorphous, oligomeric, most toxic prefibrillar phase of amyloid fibre formation characterizing neurodegenerative diseases²². Overall, our work identifies a physically interacting J-protein network that adds another level of functional flexibility to cellular protein quality control. The underlying functional basis for hitherto unexplained evolutionary maintenance of distinct J-protein classes now also becomes clear. In essence, we reveal a J-protein gearbox regulating efficacy of protein disaggregation and consequently, refolding reactions, with fundamental effect on the cellular physiology, and therefore health, of metazoan organisms.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 October 2014; accepted 8 July 2015.

Published online 5 August 2015.

- Hipp, M. S., Park, S. H. & Hartl, F. U. Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.* **24**, 506–514 (2014).

- Morimoto, R. I. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes Dev.* **22**, 1427–1438 (2008).
- Kirstein-Miles, J., Scior, A., Deuerling, E. & Morimoto, R. I. The nascent polypeptide-associated complex is a key regulator of proteostasis. *EMBO J.* **32**, 1451–1468 (2013).
- Rampelt, H. et al. Metazoan Hsp70 machines use Hsp110 to power protein disaggregation. *EMBO J.* **31**, 4221–4235 (2012).
- Goloubinoff, P., Mogk, A., Zvi, A. P., Tomoyasu, T. & Bukau, B. Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. *Proc. Natl Acad. Sci. USA* **96**, 13732–13737 (1999).
- Parsell, D. A., Kowal, A. S., Singer, M. A. & Lindquist, S. Protein disaggregation mediated by heat-shock protein HSP104. *Nature* **372**, 475–478 (1994).
- Shorter, J. The mammalian disaggregase machinery: Hsp110 synergizes with Hsp70 and Hsp40 to catalyze protein disaggregation and reactivation in a cell-free system. *PLoS ONE* **6**, e26319 (2011).
- Mayer, M. P. & Bukau, B. Hsp70 chaperones: cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* **62**, 670–684 (2005).
- Kampinga, H. H. & Craig, E. A. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nature Rev. Mol. Cell Biol.* **11**, 579–592 (2010).
- Cyr, D. M. & Ramos, C. H. Specification of Hsp70 function by type I and type II HSP40. *Subcell. Biochem.* **78**, 91–102 (2015).
- Sahi, C. & Craig, E. A. Network of general and specialty J protein chaperones of the yeast cytosol. *Proc. Natl Acad. Sci. USA* **104**, 7163–7168 (2007).
- Tzankov, S., Wong, M. J., Shi, K., Nassif, C. & Young, J. C. Functional divergence between co-chaperones of Hsc70. *J. Biol. Chem.* **283**, 27100–27109 (2008).
- Rauch, J. N. & Gestwicki, J. E. Binding of human nucleotide exchange factors to heat shock protein 70 (Hsp70) generates functionally distinct complexes *in vitro*. *J. Biol. Chem.* **289**, 1402–1414 (2014).
- Lu, Z. & Cyr, D. M. Protein folding activity of HSP70 is modified differentially by the HSP40 co-chaperones Sis1 and Ydj1. *J. Biol. Chem.* **273**, 27824–27830 (1998).
- Weibezahn, J. et al. Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cell* **119**, 653–665 (2004).
- Ramos, C. H., Oliveira, C. L., Fan, C. Y., Torriani, I. L. & Cyr, D. M. Conserved central domains control the quaternary structure of type I and type II HSP40 molecular chaperones. *J. Mol. Biol.* **383**, 155–166 (2008).
- Borges, J. C., Fischer, H., Craievich, A. F. & Ramos, C. H. Low resolution structural study of two human HSP40 chaperones in solution. DJA1 from subfamily A and DJB4 from subfamily B have different quaternary structures. *J. Biol. Chem.* **280**, 13671–13681 (2005).
- Tsai, J. & Douglas, M. G. A conserved HPD sequence of the J-domain is necessary for YDJ1 stimulation of Hsp70 ATPase activity at a site distinct from substrate binding. *J. Biol. Chem.* **271**, 9347–9354 (1996).
- Suh, W. C., Lu, C. Z. & Gross, C. A. Structural features required for the interaction of the Hsp70 molecular chaperone DnaK with its cochaperone DnaJ. *J. Biol. Chem.* **274**, 30534–30539 (1999).
- Genevaux, P., Schwager, F., Georgopoulos, C. & Kelley, W. L. Scanning mutagenesis identifies amino acid residues essential for the *in vivo* activity of the *Escherichia coli* DnaJ (Hsp40) J-domain. *Genetics* **162**, 1045–1053 (2002).
- De Los Rios, P., Ben-Zvi, A., Slutsky, O., Azem, A. & Goloubinoff, P. Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling. *Proc. Natl Acad. Sci. USA* **103**, 6166–6171 (2006).
- Caughy, B. & Lansbury, P. T. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26**, 267–298 (2003).
- Cheetham, M. E. & Caplan, A. J. Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. *Cell Stress Chaperones* **3**, 28–36 (1998).
- Li, J., Qian, X. & Sha, B. The crystal structure of the yeast Hsp40 Ydj1 complexed with its peptide substrate. *Structure* **11**, 1475–1483 (2003).
- Lu, Z. & Cyr, D. M. The conserved carboxyl terminus and zinc finger-like domain of the co-chaperone Ydj1 assist Hsp70 in protein folding. *J. Biol. Chem.* **273**, 5970–5978 (1998).

Acknowledgements We thank A. Mogk for critical reading of the manuscript and S. Ungelenk for Hsp26. This work was funded by the Deutsche Forschungsgemeinschaft (SFB1036, BU617/19-1 to B.B.; EXC257, SFB740 to J.K.), Alexander von Humboldt Foundation Postdoctoral Fellowships (to N.B.N. and A.Sz.), National Institutes of Health (the NIGMS, NIA, NIMS), Ellison Medical Foundation and Daniel F. and Ada L. Rice Foundation (to R.I.M.), German Federal Ministry of Education and Research (BMBF) Virtual Liver Network and EU FEP Flagship Programme Human Brain Project (0315749, 604102 to R.C.W.), Klaus Tschira Foundation (to M.B., A.St. and R.C.W.), Sir Henry Wellcome Postdoctoral Fellowship (to F.S.), ETH Zurich and ERC advanced grant Proteomics v3.0 (233226 to R.A.).

Author Contributions N.B.N. and B.B. conceived the study. N.B.N., J.K., A.Sz., M.B., A.St., F.S., D.L.G., R.C.W., M.P.M. and B.B. designed the experiments. N.B.N., J.K., A.Sz., M.B., A.St., F.S., K.A., X.G. and A.Sc. performed the experiments. N.B.N., J.K., A.Sz., M.B., A.St., F.S., R.A., R.C.W., R.I.M., D.L.G., M.P.M. and B.B. analysed the data. N.B.N., D.L.G., M.P.M. and B.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.B. (bukau@zmbh.uni-heidelberg.de) or N.B.N. (n.nillegoda@zmbh.uni-heidelberg.de).

METHODS

Plasmids and protein purification. Clones of human J-proteins (DNAJA1, DNAJA2, DNAJB1 and DNAJB4) were obtained either directly from Addgene or as gifts from H. Kampina in pcDNA5/FRT/TO plasmids. *C. elegans dnj-12*, *dnj-13*, C30C11.4 (*hsp-110*) and *hsp-1* genes were amplified by PCR using complementary DNA preparations from heat-shocked three-day-old animals as a template. The above-mentioned genes were then recloned into protein expression vector pCA528 or pSUMO with a 6×His-Smt3 tag as previously described²⁶. Mutants of J-proteins were generated by standard PCR mutagenesis techniques and verified by sequencing. JA2 and JB1 variants for N-terminal FLAsH and ReAsH labelling were generated by incorporating the Cys-Cys-Pro-Gly-Cys-Cys tag (5'-TGTTGTCCAGGGTGCTGC-3') after N-terminal methionine. The same labelling motif was generated in JA2 CTD by inserting two Cys residues before Pro241 and two Cys residues before Val243. The JB1 CTD labelling mutant was generated by mutating Gly278 to Cys. To obtain isolated J-domains, the J-domains of JA2 (1–77 amino acids) and JB1 (1–76 amino acids) were PCR-amplified with a C-terminal TAG site and cloned into pCA528. HPD motifs were mutated to QPN by changing H36Q+D38N and H32Q+D34N in isolated J-domains of JA2 and JB1, respectively. Purification of J-proteins and their variants was performed by affinity (Ni-IDA, Macherey-Nagel; Ni-NTA, Pierce), size-exclusion and ion-exchange chromatographic methods. In brief, BL21(D3E)/pRARE *Escherichia coli* strains carrying the corresponding expression vectors were induced for protein expression with 0.5 mM isopropyl-1-thio- β -galactopyranoside (IPTG, Sigma-Aldrich) for 3 h at 30 °C. All *C. elegans* chaperones were expressed at 20 °C with 1 mM IPTG overnight. Cells were lysed either in 50 mM HEPES-KOH, pH 7.5, 750 mM KCl, 5 mM MgCl₂, protease inhibitor cocktail (Roche), 2 mM phenylmethylsulphonyl fluoride and 10% glycerol (for human J-protein purifications), or in 30 mM HEPES-KOH, pH 7.4, 500 mM K-acetate, 5 mM MgCl₂, 1 mM β -mercaptoethanol, 2 mM phenylmethylsulphonyl fluoride, protease inhibitor cocktail (Roche) and 10% glycerol (for nematode chaperone purifications). After centrifugation at 30,000g (30 min, 4 °C) the resulting supernatants were applied to a Ni-NTA/Ni-IDA matrix and incubated for 60 min at 4 °C. Subsequent washing steps were performed with high-salt buffers (50 mM HEPES-KOH, pH 7.5, 750 and 500 mM KCl, 5 mM MgCl₂ and 10% glycerol) for human J-protein purifications. Worm chaperones were first washed in high-salt buffer (30 mM HEPES-KOH, pH 7.4, 1 M K-acetate, 5 mM MgCl₂, 1 mM β -mercaptoethanol, 2 mM phenylmethylsulphonyl fluoride and 10% glycerol), followed by a low-salt wash (identical to the high-salt buffer with 50 mM instead of 1 M K-acetate). Protein elution was performed with 300 mM imidazole in the corresponding low-salt buffers. Dialysis was performed overnight at 4 °C in the presence of 4 μ g His-tagged Ulp1 per mg substrate protein for proteolytic cleavage of the 6×His-Smt3 tag. The 6×His-Smt3 tag and His-Ulp1 were removed by incubating the dialysed proteins in Ni-NTA/Ni-IDA matrix for 60 min at 4 °C. The targeted proteins were further purified using Superdex 200 (human J-proteins and their variants), ion exchange using the Resource Q (anion exchange for DNJ-12, HSP-1, HSP-110 and isolated human J-domain fragments) or Resource 5 (cation exchange for DNJ-13) columns (GE Healthcare). Firefly luciferase and human HSPA8 and HSPH2 were purified as previously described⁴. Pyruvate kinase and α -glucosidase were purchased from Sigma-Aldrich. Pig heart muscle MDH was purchased from Roche.

Luciferase refolding and disaggregation/refolding assays. For refolding-only assays, 20 nM luciferase plus 750 nM HSPA8, 40 nM HSPH2, 380 nM J-protein and 100 nM Hsp26 was incubated at 42 °C for 10 min in HKM buffer (50 mM HEPES-KOH, pH 7.5, 50 mM KCl, 5 mM MgCl₂, 2 mM dithiothreitol (DTT), 2 mM ATP, pH 7.0 and 10 μ M BSA) to generate thermally denatured luciferase. Luciferase refolding was initiated by adding an ATP regenerating system (3 mM phosphoenolpyruvate and 20 ng μ L⁻¹ pyruvate kinase) and by shifting the reaction to 30 °C. Luciferase activity was measured at the indicated time points with a Lumat LB 9507 luminometer (Berthold Technologies) by transferring 1 μ L of sample to 100 μ L of assay buffer (25 mM glycylglycine, pH 7.4, 5 mM ATP, pH 7, 100 mM KCl and 15 mM MgCl₂) mixed with 100 μ L of 0.25 mM luciferin. Luciferase aggregates for disaggregation/refolding reactions were generated as previously described⁴. Either 25 nM or 2 μ M luciferase with fivefold excess of Hsp26 was aggregated at 45 °C for 15 min. Protein disaggregation/refolding was initiated by adding the indicated chaperone mixtures to preformed luciferase aggregates and shifting the reaction temperature to 30 °C. Luciferase disaggregation/refolding assays with *C. elegans* HSP70 chaperone system were performed at either 20 °C or 22 °C (in Fig. 1c and Extended Data Fig. 2c, d).

In cells from bacteria to human, aggregates that form under thermal stress contain small HSPs (sHSPs) such as *S. cerevisiae* Hsp26. sHSP incorporation facilitates aggregate resolution through inherent chaperone holdase activity, possibly changing the density, size and architecture of aggregates, making these more accessible and manageable for disaggregation machineries^{4,27}. The *in vitro* assay system incorporating Hsp26 therefore more closely approximates the situation in

the cell. We used *S. cerevisiae* Hsp26 because it is the only sHSP induced and activated in yeast cells after heat shock, and has been extensively characterized *in vivo* and *in vitro*^{4,27,28}. Hsp26 is therefore the generic heat-induced sHSP of yeast, which justifies its use for our study. By contrast, there are 10 human sHSPs, each with different substrate binding specificities and affinities^{29,30}, and there is no clear basis for choosing one above another. Also, some of these, including those recognizing luciferase as substrate, interact to form hetero-oligomers, which display yet further different properties³¹ resulting in complicated assay analysis. Furthermore, the activation mechanism of at least one of these is subject to controversy (phosphorylation and dephosphorylation have both been reported³²; P. Goloubinoff, personal communication).

α -glucosidase disaggregation assay. α -glucosidase aggregation was achieved by incubating 50 nM of substrate with 500 nM Hsp26 in HKM buffer without DTT at 50 °C for 15 min. Disaggregation/refolding of aggregates was initiated by adding indicated chaperone mixtures supplemented with an ATP regenerating system and by shifting the reaction to 30 °C. Reactivation of α -glucosidase was measured with an α -glucosidase assay kit from Abnova using a FLUOstar Omega plate-reader (BMG LABTECH).

MDH disaggregation assay. Pig heart muscle MDH (Roche) aggregation was achieved by incubating 150 nM of substrate with 750 nM Hsp26 in HKM buffer at 47 °C for 30 min. Disaggregation/refolding of aggregates was initiated by adding indicated chaperone mixtures supplemented with an ATP regenerating system and by shifting the reaction to 30 °C. MDH activity was measured using a potassium phosphate (150 mM, pH 7.6) buffer containing 1 mM oxaloacetate, 2 mM DTT and 0.56 mM NADH. Activity measurements were taken using a FLUOstar Omega plate-reader. Refolding rates were calculated from the linear increase of substrate activities.

SEC and aggregate profiling. Tritium (³H) labelling of firefly luciferase, α -glucosidase, JA2 and JB1 was performed with *N*-succinimidyl-[2,3-³H]propionate (Hartmann Analytic) according to manufacturer's guidelines. Unincorporated *N*-succinimidyl-[2,3-³H]propionate was removed using dialysis with HKM buffer and either 150 mM (for luciferase and α -glucosidase) or 500 mM (for J-proteins) plus KCl at 4 °C, overnight. ³H-labelled luciferase and α -glucosidase were aggregated as described in the disaggregation/refolding assays. Luciferase and α -glucosidase reactivation were performed by adding specified chaperone cocktails and incubating at 30 °C. Reactions were quenched with apyrase (0.8 μ g μ L⁻¹) at 40 or 120 min and placed on ice. Aggregated luciferase/ α -glucosidase complexes were separated using an ÄKTA purifier system with a Superose 6 Tricorn column (10/300 GL, GE Healthcare Life Sciences). Samples were centrifuged at 9,000g for 5 min at 4 °C before loading. Running buffer contained 50 mM HEPES-KOH, pH 7.5, 150 mM KCl, 5 mM MgCl₂, 2 mM DTT (0.2 mM DTT for α -glucosidase) and 10% glycerol. A similar approach was used to separate ³H-labelled J-protein dimers, with the exception of using a buffer with 50 mM KCl. The ³H signal in each fraction (500 μ L) was quantified by scintillation counting (Beckman LS6000 IC). The amount of ³H-luciferase trapped in GroEL^{D87K} was calculated by subtracting the total counts between elution volumes 11 and 16 in reactions without the trap from that of the reaction containing the trap (10 μ M). The ³H signal in each elution fraction was normalized to the total counts of the corresponding SEC run after background subtraction, and presented as a percentage of the total counts (F1–F4). A SEC standard (BIO-RAD) was used to determine the size of the elution peaks. Void volume contains any complexes \geq 5,000 kDa. Notably, ~40–50% of the input material was lost during SEC as a result of nonspecific binding to filters and column matrix.

Luciferase aggregation prevention assay. In the aggregation prevention assay, 200 nM luciferase was mixed with indicated concentrations of chaperones or BSA (control) in a buffer containing 50 mM HEPES-KOH, pH 7.5, 50 mM KCl, 5 mM MgCl₂, 2 mM DTT and 2 mM ATP, pH 7.0. Aggregation of luciferase was initiated by increasing the temperature to 42 °C. The extent of luciferase aggregation was monitored by light scattering at 600 nm (Hitachi Fluorometer F4500, $\lambda_{ex/em}$ = 600 nm, slit widths of 5.0 nm) for 25 or 30 min.

Chemical crosslinking coupled to mass spectrometry. For chemical cross-linking, 100 μ L of sample containing 2 μ M JA2 and 2 μ M JB1 was directly cross-linked with 1 mM disuccinimidyl suberate d0/d12 (DSS, Creativemolecules Inc.), and subsequently enzymatically digested with trypsin and enriched for cross-linked peptides, essentially as previously described³³. Liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis was carried out on an Orbitrap Elite mass spectrometer (Thermo Electron). Data were searched using xQuest³⁴ in iontag mode with a precursor mass tolerance of 10 p.p.m. For matching of fragment ions, tolerances of 0.2 Da for common ions and 0.3 Da for cross-link ions were used. False discovery rates of cross-linked peptides were assigned using xProphet³⁵. Cross-linked peptides were identified with a delta score <0.95 and a linear discriminant score >20, and additionally analysed by visual inspection to

ensure good matches of ion series on both cross-linked peptide chains for the most abundant peaks.

FLAsH and ReAsH labelling of J-proteins. J-protein variants with introduced tetracysteine motif (CCPGCC, for FLAsH and ReAsH labelling) or with single cysteine residue (for JB1 Alexa Fluor 488 labelling) were reduced with 50-fold molar excess of TCEP for 30 min at room temperature, and incubated with FLAsH or ReAsH (gift from A. Krezel) at a 1:1.5 protein/label ratio for 4 h at 4 °C, or 20-fold excess of Alexa-Fluor-488-maleimide for 2 h at room temperature. Progress of labelling reaction for the biarsenical dyes was monitored in a spectrophotometer, using $A_{280\text{ nm}}$ (JA2 $\epsilon_{280\text{ nm}} = 24,000\text{ M}^{-1}\text{ cm}^{-1}$, JB1 $\epsilon_{280\text{ nm}} = 19,035\text{ M}^{-1}\text{ cm}^{-1}$) for protein concentration and absorption at $A_{510\text{ nm}}$ for FLAsH ($\epsilon_{510\text{ nm}} = 41,000\text{ M}^{-1}\text{ cm}^{-1}$), $A_{590\text{ nm}}$ for ReAsH ($\epsilon_{590\text{ nm}} = 68,000\text{ M}^{-1}\text{ cm}^{-1}$) and $A_{494\text{ nm}}$ for Alexa Fluor 488 ($\epsilon_{494\text{ nm}} = 71,000\text{ M}^{-1}\text{ cm}^{-1}$). The excess of unbound dye was removed on a self-packed Sephadex G-25 column (GE Healthcare), and the activity of labelled proteins was confirmed by luciferase refolding assay.

FRET measurements. FRET was used to validate distances between labelled J-domains and CTDs of class A and B J-proteins. Emission spectra were recorded on a Jasco FP6500 spectrofluorimeter between 510 and 650 nm, at excitation wavelength of 508 nm for FLAsH and 488 nm for Alexa Fluor 488 (donor fluorophores). Quenching of donor fluorescence (at 519 nm for Alexa Fluor 488 and 533 nm for FLAsH) and an increase in acceptor emission (at 608 nm for ReAsH) were quantified. Acceptor fluorescence measurement was refined by subtracting the fluorescence from donor-labelled J-protein to minimize background. The Förster radius of the FLAsH–ReAsH FRET pair was calculated to 39 Å (refs 36, 37) and Alexa-Fluor-488–ReAsH to 62 Å (ref. 38). For FRET experiments, J-proteins labelled with donor and acceptor fluorophores were mixed at 0.1 µM (donor) and 1 µM (acceptor) in a buffer containing 25 mM HEPES, pH 7.5, 50 mM KCl and 5 mM MgCl_2 , and allowed to equilibrate for 15 min at 30 °C before measuring the steady-state fluorescence. For competition experiments, 1-, 5- and 10-fold excess (relative to acceptor concentration) of unlabelled full-length proteins or isolated J-domains were added and allowed to equilibrate for 15 min at 30 °C. All samples were measured at least in duplicate. For Figs 2b and 3d, FRET efficiencies were calculated based on the donor fluorescence quenching, and presented as a percentage of donor fluorescence in the absence of acceptor.

Protein structure preparation. Protein structures used in simulations were either crystal or NMR structures from the RCSB Protein Data Bank (PDB; <http://www.rcsb.org>) or comparative models that were either present in the SWISS-MODEL database and found using the Protein Model Portal (PMP) or were modelled with SWISS-MODEL (SM)^{39–42}. The structure of the CTD^{JB1} dimer was taken from the crystal structure (PDB code 3AGZ, resolution: 2.51 Å)⁴³ and that of JD^{JB1} from the first entry of the NMR structure (PDB code 1HDJ)⁴⁴. Since the N-terminus of chain B in the CTD^{JB1} dimer was missing three residues compared to chain A, the N-terminal nine residues from chain A were superimposed on the N-terminus of chain B to obtain coordinates for the missing three residues. Comparative models of the CTD^{JB4} dimer and JD^{JB4} were both found in the PMP and were based on the template structures 3AGZ and 1HDJ, respectively. To add the missing three residues at the N-terminus of chain B, the same procedure as for CTD^{JB1} was used. Comparative models of CTD monomers of JA1 and JA2 were both taken from the PMP. Both structures were modelled with SM based on the template crystal structure, 1NLT (resolution 2.70 Å)²⁴. The structure of JD^{JA1} was the first entry of the NMR structure, 2LO1 in the RCSB PDB. A comparative model of JD^{JA2} was taken from the PMP and was based on the template structure, 2LO1. Structures of the CTD dimer were generated for JA1 and JA2 as follows: the dimerization site was modelled with SM based on the template crystal structure, 1XAO (resolution 2.07 Å). Then, the structures of the CTD monomers were superimposed on the corresponding dimerization site model and only C-terminal missing residues of the dimerization site were added to the CTD domains. The structure of the J-domain of DNJ-12 was taken from the crystal structure PDB code 2OCH (resolution 1.86 Å). The CTD dimer of DNJ-12 was modelled based on the crystal structure, 1NLT, using SM. The J-domain of DNJ-13 was modelled based on 1HDJ and the dimer structure of the CTD of DNJ-13 was based on 3AGZ A/B. Further editing of the following structures was performed to generate a set of comparable structures of the J-proteins. N-terminal Gly was deleted in JD^{JA1}, because it was not part of the UniProt entry P31689. The last seven residues of JD^{JB1} were deleted to have a comparable C-terminal end to the J-domains of JA1 and JA2. Similarly, the last four C-terminal residues in JD^{JB4} were deleted to obtain comparable C-terminal ends to the J-domains of JA1 and JA2.

Protein–protein docking. Protein–protein docking was performed with a rigid-body treatment of the protein structures using the Simulation of Diffusional Association (SDA) program (version 7, <http://mcm.h-its.org/sda7>)^{45,46}. SDA uses Brownian dynamics (BD) simulation to perform the sampling of protein configurations subject to inter-protein forces and torques due to electrostatic and non-polar interactions. The docking protocol required the following steps:

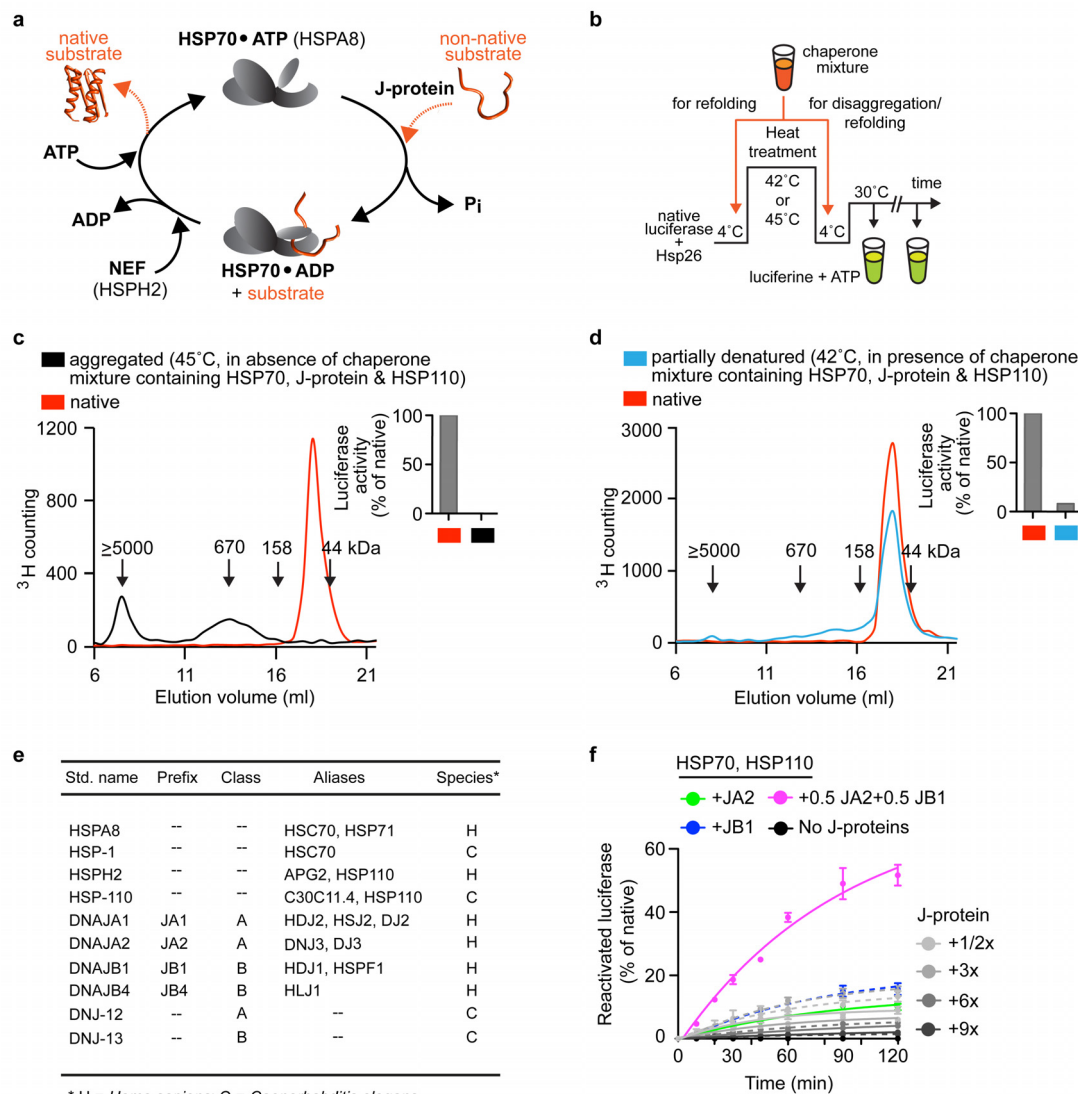
1. Structure preparation: polar hydrogen atoms were added to the protein structures with WHATIF5 (ref. 47) assuming pH 7.2.
2. Calculation of protein electrostatic potentials: the electrostatic potential of each protein structure was calculated by numerically solving the linearized Poisson–Boltzmann equation with UHBD (ref. 48). Electrostatic potential grids with 250^3 grid points with a 1 Å spacing were used for all proteins. The dielectric constants of the solvent and the protein were set to 78.0 and 4.0, respectively, and the dielectric boundary was defined by the protein's van der Waals surface. The ionic strength was set to 50 mM at a temperature of 300K, with an ion exclusion radius (Stern layer) of 1.5 Å. The protein atoms were assigned OPLS atomic partial charges and radii⁴⁹.
3. Calculation of effective charges: these were derived with ECM⁵⁰. The effective charges for each protein were fit to reproduce the electrostatic potential in a 3-Å-thick layer extending outwards from the protein's solvent-accessible surface computed as defined by a probe of radius 4 Å. The effective charges for proteins were placed on the carboxylate oxygen atoms of Asp and Glu amino acid residues and the C-terminus, and the amine nitrogen atoms of Lys and Arg amino acid residues and the N-terminus. For the Zn^{2+} ion, an effective charge site with a formal charge of $-2e$ was placed on the ion, corresponding to the summed charge of the ion and its four coordinating cysteine side-chains.
4. Calculation of polar desolvation grids: the desolvation penalty of each effective charge was computed as the sum of desolvation penalties due to the low dielectric cavity of each atom of the other protein⁵¹, which was precomputed on a grid. The grid dimensions were set to 150^3 grid points with a spacing of 1 Å. Ionic strength and dielectric constants were assigned as for the electrostatic potential calculations. The ion radius was assigned as 1.5 Å.
5. Calculation of non-polar desolvation grids: the non-polar desolvation forces were computed using precomputed grids⁵². The distance parameters (a) and (b) were assigned values of 3.10 Å and 4.35 Å, respectively. The parameter (c) was assigned as 1.0 and the conversion factor to $\beta = -0.0065\text{ kcal mol}^{-1}\text{ Å}^{-2}$. The grid dimensions were set to 150^3 grid points with a spacing of 1 Å.
6. Calculation of excluded volume grids: protein shape was described by a grid with a 0.25 Å spacing. A probe of radius of 1.77 Å was used to determine the protein shape. The radius of the solvent probe to determine surface atoms was set to 1.4 Å.
7. Docking simulations: for each protein pair docked, 10,000 trajectories were generated with SDA. Trajectories were started with the proteins at a separation distance of 100 Å and a random relative orientation. A trajectory was terminated if the protein separation exceeded 300 Å or a simulation time of 500 ns was reached. The protein–protein separation was calculated as the distance between their centres of geometry (CoG). Up to 3,000 configurations sampled during the BD trajectories with a separation of less than 105 Å were saved. During the BD simulations, if a new docking pose was considered similar to a previously saved pose, that is, had an approximate root mean squared deviation (r.m.s.d.) less than 2 Å, then the configuration with the lower intermolecular energy was saved and the counter of this docking pose, the occupation, was incremented. The relative translational diffusion coefficient was set to $0.027\text{ Å}^2\text{ ps}^{-1}$. The rotational diffusion coefficient for both proteins was set to $3.92 \times 10^4\text{ radian}^2\text{ ps}^{-1}$. The time step was 1 ps at separations less than 120 Å and increased linearly beyond this threshold with slope of 2 ps Å^{-1} .
8. Clustering: for each protein pair docked, the configurations saved were clustered with a hierarchical clustering algorithm. The backbone r.m.s.d. between each docked protein configuration was calculated to produce an inter-configuration distance matrix. Initially, each docked structure was assigned to a separate cluster. The closest clusters were found and merged; the distance matrix was updated. This process was repeated until all docked protein structures were in one cluster. The distance between clusters was defined as the average backbone r.m.s.d. between docked protein structures in one cluster relative to structures in another cluster. The representative of a cluster is the protein configuration with the smallest r.m.s.d. to every other member of the cluster. In each clustering cycle, the mean and s.d. of the r.m.s.d. of all members of each cluster to the corresponding cluster representative were calculated. The number of configurations in each cluster in each clustering cycle was determined, taking account of cluster occupation during the BD simulations, and the clusters were ranked by size. The number of generated clusters was chosen using the following criteria. Starting with the largest cluster, the minimum number of clusters accounting for 90% of the total number of configurations docked and satisfying the criterion that the mean r.m.s.d. plus s.d. of the clusters is less than 10 Å, was

determined. This threshold results in configurations with similar CoG of the J-domains but orientations differing by about 90° being assigned to different clusters.

Of note, class B CTD dimers mostly showed a higher number of selected and total clusters with less favourable interaction energies than class A CTD dimers. This indicates a higher diversity in the docking poses of J-domains to class B CTD dimers (see clustering at N-termini of class B CTD dimers, Extended Data Fig. 7). However, in a full-length protein, this clustering region would be occupied by glycine-phenylalanine-rich linkers and J-domains as in the crystal structure of class B DnaJ2 from *Thermus thermophilus* (PDB code 4J80)⁵³. Consistently, a docking simulation performed with full-length DnaJ2 showed a much more specific interaction of the DnaJ2 J-domain centering between CTD-I and CTD-II regions (data not shown). To take this into account, we also analyzed the docking clusters for JD^{JA2} and CTD^{JB1} by requiring docking positions to satisfy distance requirements from cross-linking data (see Fig. 2d and Extended Data Fig. 7).

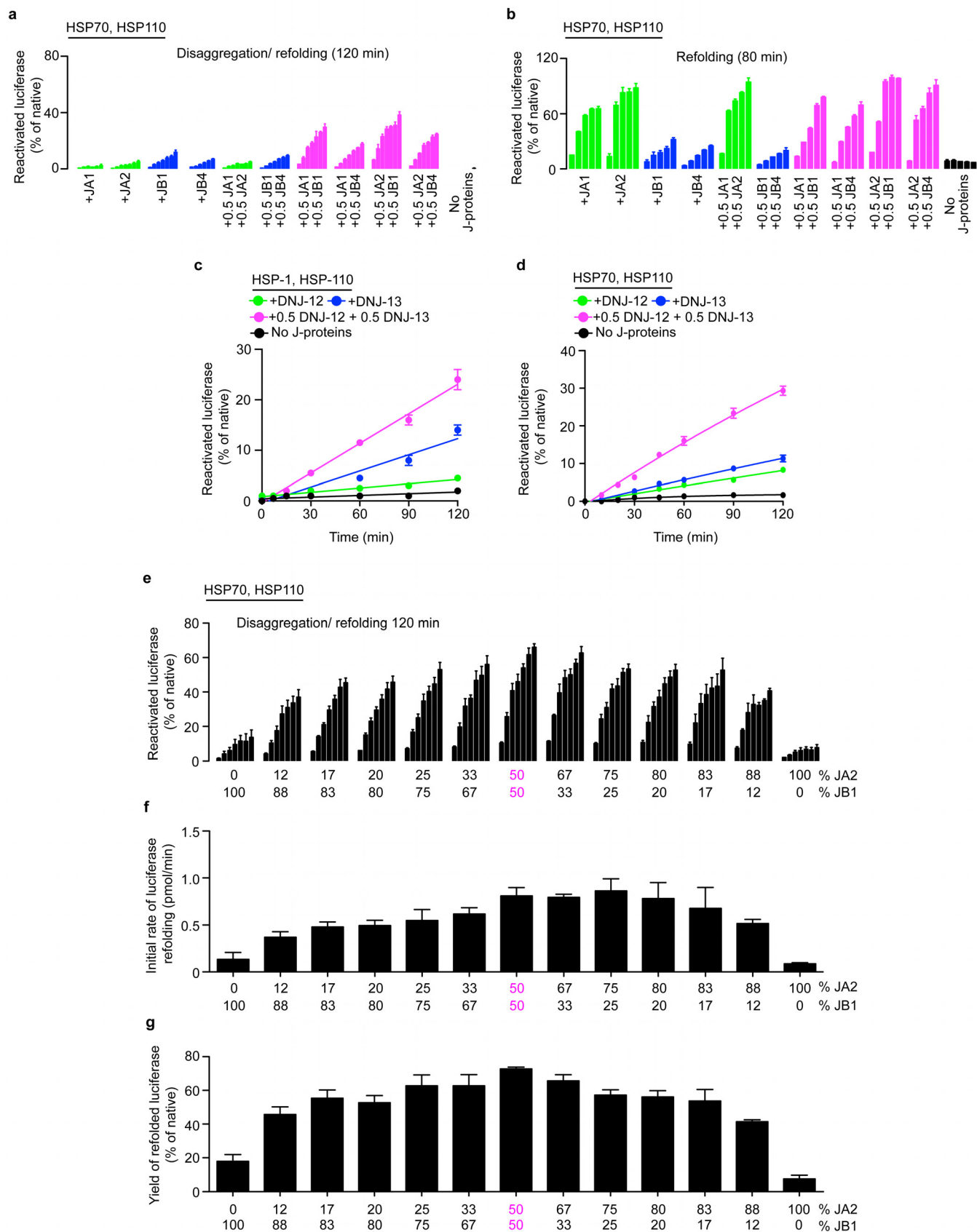
Modelling of JA2 and JB1 complexes. Using PyMOL (<http://www.pymol.org>) software, two putative arrangements (compact and open) of the JA2 and JB1 CTD homodimers were generated to satisfy the maximum number of observed cross-links and FRET constraints possible. The J-domains of JA2 and JB1 were added ensuring consistency with docking, cross-linking and FRET results and a distance to the corresponding CTD that would be allowed by the missing residues that connect J-domain and CTD. The positions of the J-domains of JB1 and JA2 in the two models are expected to provide a FRET efficiency of 8 and 3% (within experimental error), correlating with the lack of experimental observation of FRET between these domains. Note that the structures were treated as rigid bodies and flexibility of the CTD dimers parallel and perpendicular to the dimer plane is very likely and would allow other configurations of these complexes.

26. Andréasson, C., Fiaux, J., Rampelt, H., Mayer, M. P. & Bukau, B. Hsp110 is a nucleotide-activated exchange factor for Hsp70. *J. Biol. Chem.* **283**, 8877–8884 (2008).
27. Cashikar, A. G., Duennwald, M. & Lindquist, S. L. A chaperone pathway in protein disaggregation. Hsp26 alters the nature of protein aggregates to facilitate reactivation by Hsp104. *J. Biol. Chem.* **280**, 23869–23875 (2005).
28. Haslbeck, M., Miess, A., Stromer, T., Walter, S. & Buchner, J. Disassembling protein aggregates in the yeast cytosol. The cooperation of Hsp26 with Ssa1 and Hsp104. *J. Biol. Chem.* **280**, 23861–23868 (2005).
29. Carra, S. *et al.* Different anti-aggregation and pro-degradative functions of the members of the mammalian sHSP family in neurological disorders. *Phil. Trans. R. Soc. Lond. B* **368**, 20110409 (2013).
30. Vos, M. J. *et al.* HSPB7 is the most potent polyQ aggregation suppressor within the HSPB family of molecular chaperones. *Hum. Mol. Genet.* **19**, 4677–4693 (2010).
31. Skouri-Panet, F., Michiel, M., Ferard, C., Duprat, E. & Finet, S. Structural and functional specificity of small heat shock protein HspB1 and HspB4, two cellular partners of HspB5: role of the *in vitro* hetero-complex formation in chaperone activity. *Biochimie* **94**, 975–984 (2012).
32. Peschek, J. *et al.* Regulated structural transitions unleash the chaperone activity of α B-crystallin. *Proc. Natl Acad. Sci. USA* **110**, E3780–E3789 (2013).
33. Leitner, A. *et al.* Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol. Cell. Proteomics* **11**, M111.014126 (2012).
34. Rinner, O. *et al.* Identification of cross-linked peptides from large sequence databases. *Nature Methods* **5**, 315–318 (2008).
35. Walzthoeni, T. *et al.* False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nature Methods* **9**, 901–903 (2012).
36. Adams, S. R. *et al.* New biarsenical ligands and tetracysteine motifs for protein labeling *in vitro* and *in vivo*: synthesis and biological applications. *J. Am. Chem. Soc.* **124**, 6063–6076 (2002).
37. Spagnuolo, C. C., Vermeij, R. J. & Jares-Erijman, E. A. Improved photostable FRET-competent biarsenical-tetracysteine probes based on fluorinated fluoresceins. *J. Am. Chem. Soc.* **128**, 12040–12041 (2006).
38. Dixit, A., Ray, K., Lakowicz, J. R. & Black, L. W. Dynamics of the T4 bacteriophage DNA packaging motor: endonuclease VII resolvase release of arrested Y-DNA substrates. *J. Biol. Chem.* **286**, 18878–18889 (2011).
39. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30** (suppl. 1), S162–S173 (2009).
40. Kopp, J. & Schwede, T. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res.* **32**, D230–D234 (2004).
41. Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L. & Schwede, T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* **37**, D387–D392 (2009).
42. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
43. Suzuki, H. *et al.* Peptide-binding sites as revealed by the crystal structures of the human Hsp40 Hdj1 C-terminal domain in complex with the octapeptide from human Hsp70. *Biochemistry* **49**, 8577–8584 (2010).
44. Qian, Y. Q., Patel, D., Hartl, F. U. & McColl, D. J. Nuclear magnetic resonance solution structure of the human Hsp40 (Hdj1) J-domain. *J. Mol. Biol.* **260**, 224–235 (1996).
45. Martinez, M. *et al.* SDA7: a modular and parallel implementation of the simulation of diffusional association software. *J. Comput. Chem.* **36**, 1631–1645 (2015).
46. Gabdoulline, R. R. & Wade, R. C. Simulation of the diffusional association of barnase and barstar. *Biophys. J.* **72**, 1917–1929 (1997).
47. Vriend, G. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–56 (1990).
48. Madura, J. D. *et al.* Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics Program. *Comp. Phys. Comm.* **91**, 57–95 (1995).
49. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
50. Gabdoulline, R. R. & Wade, R. C. Effective charges for macromolecules in solvent. *J. Phys. Chem.* **100**, 3868–3878 (1996).
51. Elcock, A. H., Gabdoulline, R. R., Wade, R. C. & McCammon, J. A. Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin. *J. Mol. Biol.* **291**, 149–162 (1999).
52. Gabdoulline, R. R. & Wade, R. C. On the contributions of diffusion and thermal activation to electron transfer between *Phormidium laminosum* plastocyanin and cytochrome *f*: Brownian dynamics simulations with explicit modeling of nonpolar desolvation interactions and electron transfer events. *J. Am. Chem. Soc.* **131**, 9230–9238 (2009).
53. Barends, T. R. *et al.* Combining crystallography and EPR: crystal and solution structures of the multidomain cochaperone DnaJ. *Acta Crystallogr. D* **69**, 1540–1552 (2013).
54. Dragovic, Z., Broadley, S. A., Shomura, Y., Bracher, A. & Hartl, F. U. Molecular chaperones of the Hsp110 family act as nucleotide exchange factors of Hsp70s. *EMBO J.* **25**, 2519–2528 (2006).
55. Raviol, H., Sadlish, H., Rodriguez, F., Mayer, M. P. & Bukau, B. Chaperone network in the yeast cytosol: Hsp110 is revealed as an Hsp70 nucleotide exchange factor. *EMBO J.* **25**, 2510–2518 (2006).
56. Chen, J., Walter, S., Horwich, A. L. & Smith, D. L. Folding of malate dehydrogenase inside the GroEL-GroES cavity. *Nature Struct. Biol.* **8**, 721–728 (2001).
57. Linke, K., Wolfram, T., Bussemer, J. & Jakob, U. The roles of the two zinc binding sites in DnaJ. *J. Biol. Chem.* **278**, 44457–44466 (2003).
58. Garimella, R. *et al.* Hsc70 contacts helix III of the J domain from polyomavirus T antigens: addressing a dilemma in the chaperone hypothesis of how they release E2F from pRb. *Biochemistry* **45**, 6917–6929 (2006).
59. Greene, M. K. *et al.* Role of the J-domain in the cooperation of Hsp40 with Hsp70. *Proc. Natl Acad. Sci. USA* **95**, 6108–6113 (1998).



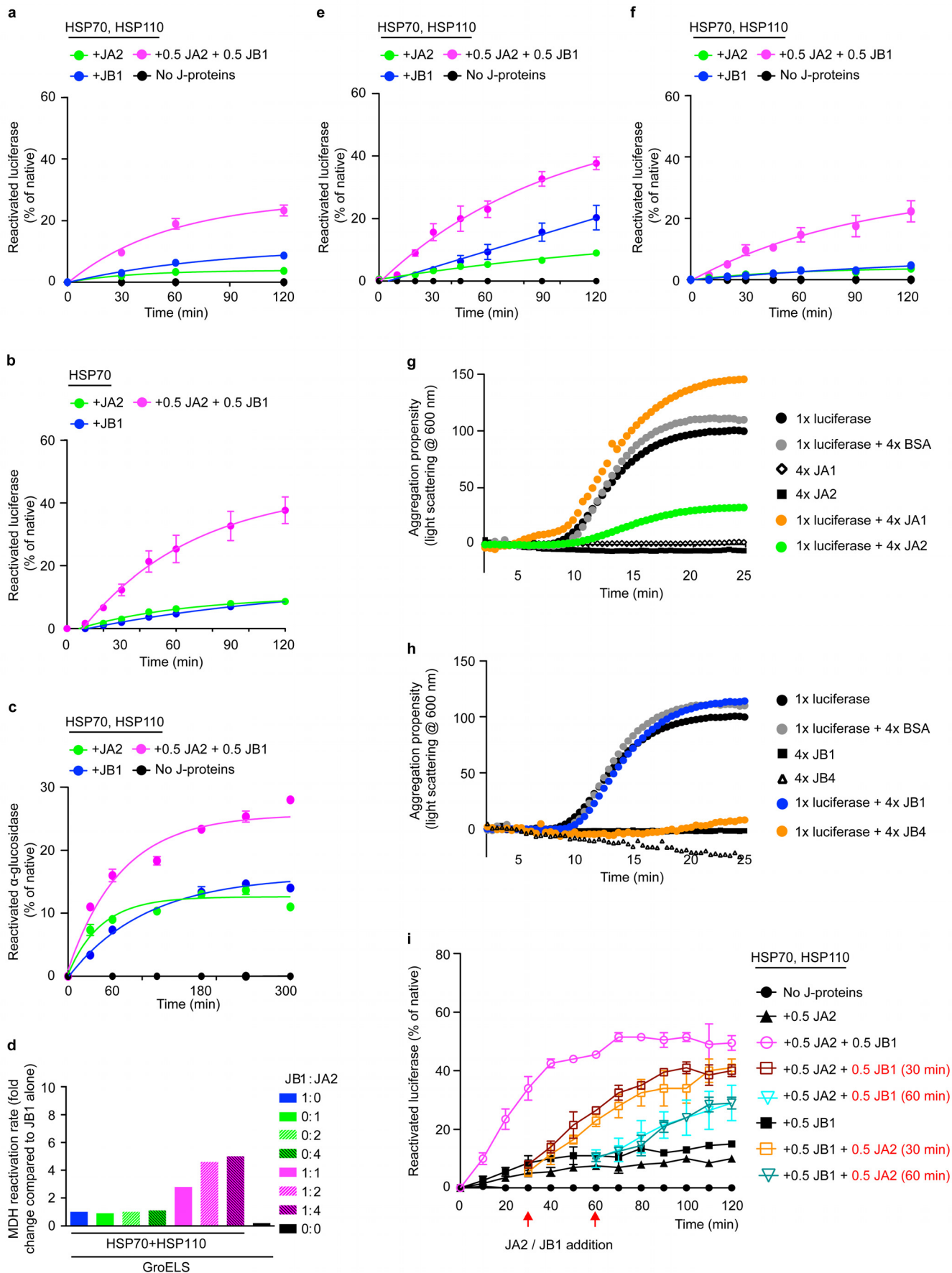
Extended Data Figure 1 | Characterization of protein disaggregation/refolding and refolding-only reactions. **a**, HSP70–J-protein–HSP110 (HSPA8–J-protein–HSPH2) functional cycle. Concomitant interaction of HSP70 with a J-protein and substrate results in allosteric stimulation of ATP hydrolysis; this traps the substrate in HSP70 (ref. 8). Subsequent NEF (for example, HSP110) promoted ADP dissociation from HSP70, then allows ATP rebinding, which triggers substrate release to complete the cycle^{54,55}. **b**, Scheme for *in vitro* disaggregation/refolding and refolding-only reactions. The aggregates used in disaggregation/refolding assays are preformed by heating luciferase with yeast small heat-shock protein (sHSP) Hsp26 (ref. 4), which is known to co-aggregate with misfolded proteins *in vivo*^{27,28} (see Methods for detailed description). If HSP70, J-protein and HSP110 are instead heated together with substrate and Hsp26, luciferase is denatured into a more easily

refoldable, inactive and largely monomeric substrate form used in refolding-only assays. **c**, SEC profiles of aggregated ³H-labelled luciferase (black; size range 200 kDa to ≥5,000 kDa representing ~2 to >50 aggregated luciferase molecules) and monomeric native luciferase (red; size ~63 kDa). Arrows indicate elution size (kDa). Inset, activity of loaded material. **d**, SEC profile of partially denatured and largely monomeric luciferase (starting material for refolding-only reactions). Inset, activity of loaded material. **e**, Chaperone nomenclature. **f**, Disaggregation and reactivation of preformed luciferase aggregates using human HSP70–HSP110 with human J-proteins JA2, green; JB1, blue; JA2+JB1, magenta or no J-protein, black. Under limiting chaperone (HSP70/HSP110) and increasing J-protein concentrations (A, solid or B, dashed) (*n* = 3). Data are mean ± s.e.m. Precise concentrations are shown in Extended Data Table 1.



Extended Data Figure 2 | Effects of mixed-class J-proteins on disaggregation/refolding and refolding-only activity of the HSP70 system. **a**, Disaggregation/refolding of aggregated luciferase compared for human class A (JA1 and JA2) and class B (JB1 and JB4) J-proteins ($n = 3$). **b**, Luciferase refolding-only compared for JA1, JA2, JB1 and JB4 ($n = 3$). **c**, Reactivation of heat-aggregated luciferase with nematode HSP70 machinery, using reduced substrate:HSP70 ratio of 1:20, containing DNJ-12 (A), DNJ-13 (B) or

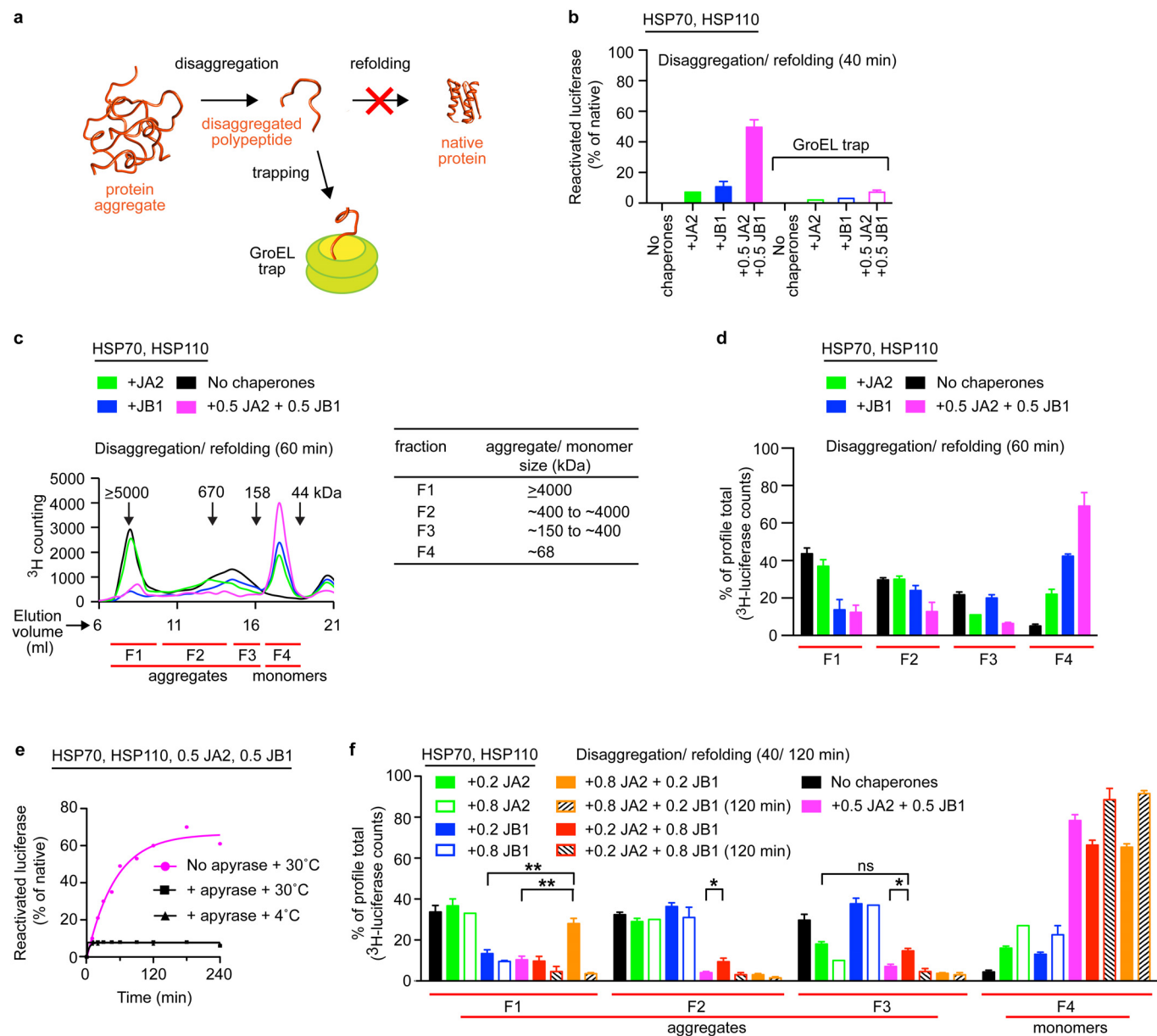
DNJ-12 + DNJ-13 (A + B) ($n = 2$). **d**, Disaggregation/refolding of luciferase using human HSP70 and HSP110 combined with nematode J-proteins ($n = 3$). **e**, Reactivation of luciferase showing optimal JA2:JB1 ratio for disaggregation/refolding ($n = 2$). **f**, Initial disaggregation/refolding rates for **e**. **g**, Final yields of refolded luciferase (120 min) for **e**. Data are mean \pm s.e.m. Precise concentrations are shown in Extended Data Table 1.



Extended Data Figure 3 | Disaggregation synergy is independent of sHSP incorporation, NEF, substrate and aggregate character, and is not explained by sequential J-protein class activity. **a**, Disaggregation/refolding reaction for luciferase aggregates without incorporating sHSP Hsp26 ($n = 3$).

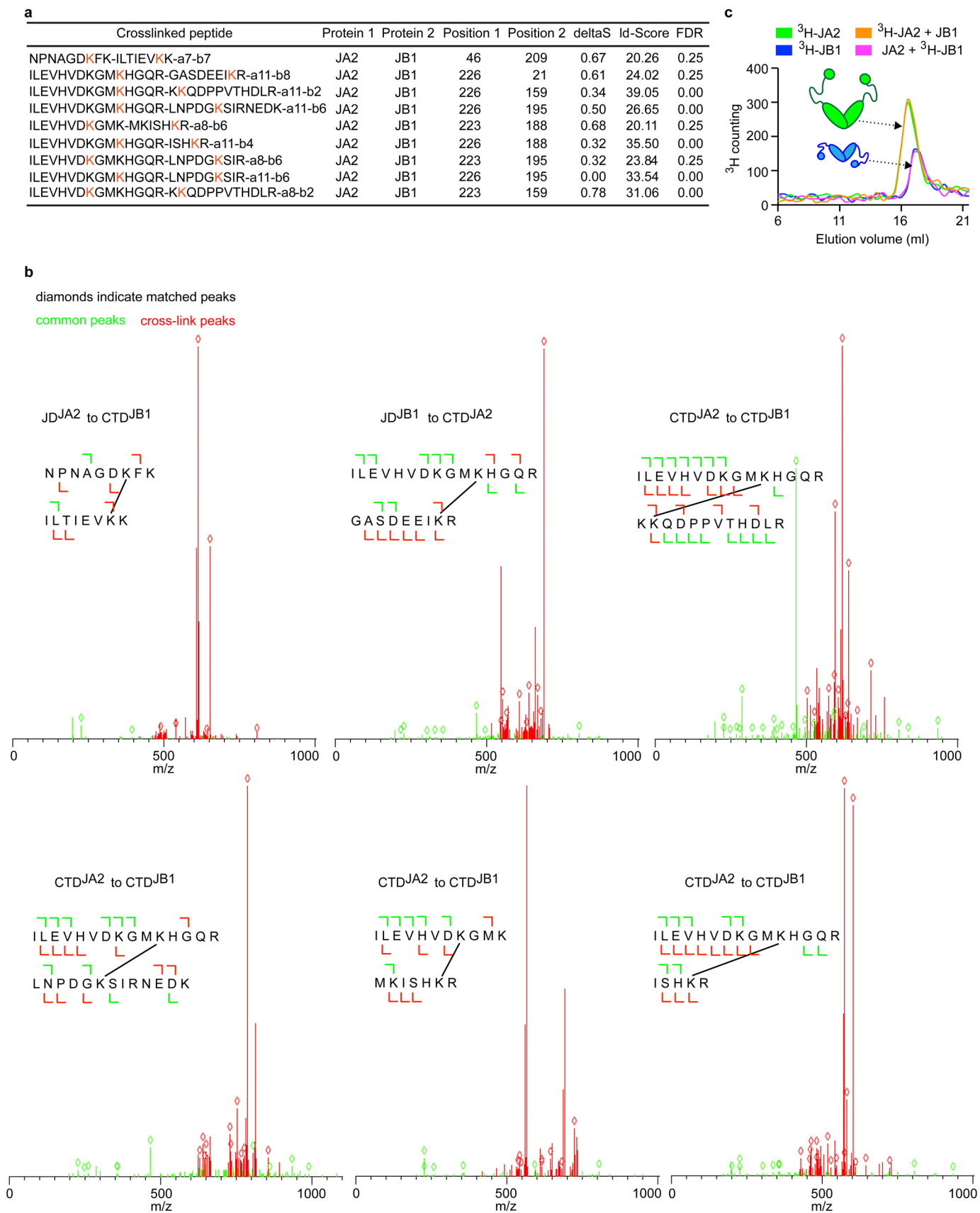
b, Reactivation without NEF (HSPH2) ($n = 3$). **c**, Reactivation of α -glucosidase aggregates ($n = 3$). **d**, Reactivation of preformed MDH aggregates in the presence of GroEL plus the GroES protein foldase system (GroELS) ($n = 2$). GroELS is required for efficient MDH refolding⁵⁶. GroELS alone is in black. JB1:JA2 denotes the stoichiometry of each reaction. **e**, Disaggregation/refolding of stringent aggregates ($\geq 5,000$ kDa) formed using 2 μ M luciferase ($n = 3$).

f, Disaggregation/refolding of aggregated luciferase at reduced substrate:HSP70 ratio (luciferase:HSP70:J-protein:HSP110 = 1:7.5:3.8:0.4) ($n = 3$). The aggregated luciferase concentration is 100 nM. **g**, **h**, Holdase function of J-proteins (class A (**g**) and class B (**h**)) during luciferase aggregation at 42 °C, shown by decreased light scattering. Concentrations: 1 \times luciferase; 4 \times J-protein; 4 \times BSA (control) ($n = 2$). **i**, Reactivation with sequential JA2 and JB1 addition. J-protein added at $t = 0$ min (black graph legends); J-protein added after 30 or 60 min (red graph legends and arrows) ($n = 2$). Data are mean \pm s.e.m. Precise concentrations are shown in Extended Data Table 1.



Extended Data Figure 4 | Stoichiometry of class A and B J-proteins determines the range of aggregate sizes resolved. **a**, The GroEL^{D87K} facilitates the capture of ³H-luciferase monomers liberated by protein disaggregation before the refolding step. **b**, Refolding of disaggregated ³H-luciferase monomers (40 min) in the absence (solid bars) and presence of GroEL trap (open bars). **c**, SEC profile after disaggregation/refolding of aggregated tritiated α -glucosidase (60 min) with either J-protein class alone (green (A) or blue (B)) or J-proteins combined (magenta). Control reaction without chaperones (black). Elution fractions F1–F4 (red lines). Table shows size distribution of aggregates in each fraction; F1 luciferase aggregates $\geq 4,000$ kDa; F2, aggregates ~ 400 – $4,000$ kDa; F3, aggregates ~ 150 – 400 kDa, F4 disaggregated monomers (~ 68 kDa). **d**, Quantification of SEC profile

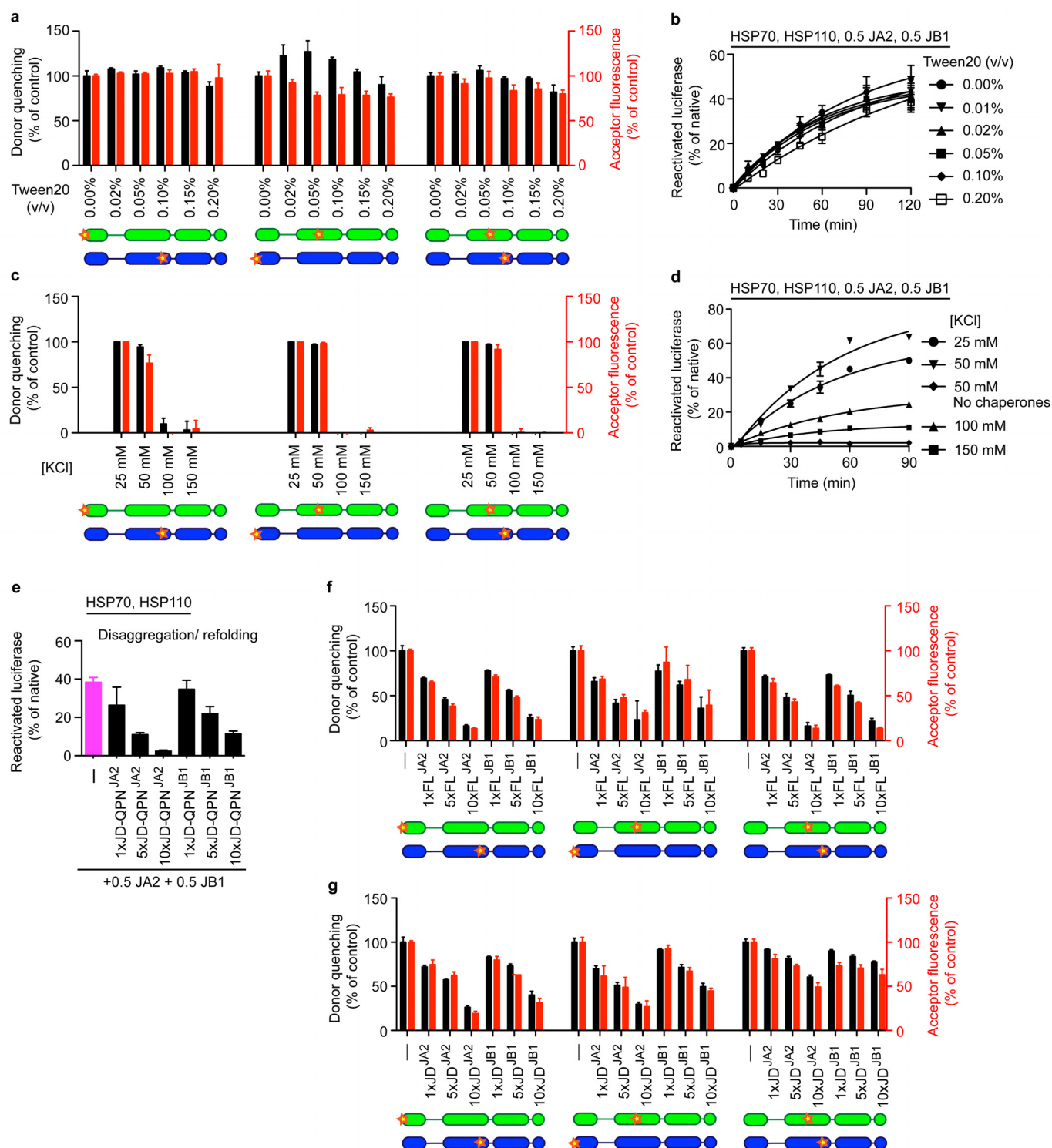
measuring disaggregation of tritiated α -glucosidase from aggregates (F1–F3) from **c**, also showing concomitant accumulation of disaggregated monomer (F4) from **c** ($n = 3$). **e**, ATP depletion by apyrase abrogates disaggregation. **f**, Quantification of SEC profile measuring disaggregation of tritiated luciferase from aggregates (F1–F3) with concomitant accumulation of disaggregated monomer (F4), using the HSP70–HSP110 system with JA2 or JB1 alone, or with JA2 plus JB1. Stoichiometry range used for JA2:JB1, 1:1 to 4:1 to 1:4. Specifically, 0.2 JB1:0.8 JA2 (orange); 0.2 JA2:0.8 JB1 (red). Solid colours denote 40-min reaction time; hash denotes 120 min. Control reaction without chaperones (black). Two-tailed t -test, * $P < 0.05$, ** $P < 0.01$ ($n = 3$). Data are mean \pm s.e.m. Precise concentrations are shown in Extended Data Table 1.



Extended Data Figure 5 | JA2 and JB1 form homodimers and interact

transiently. **a**, Identified JA2 and JB1 inter-molecular cross-links; 'Id', amino acid sequence of peptides showing cross-linked lysines (K, orange). Protein 1 and 2 denote source proteins for cross-linked peptides; position 1 and 2 denote positions of cross-linked lysines within proteins; deltaS is the delta score for each crosslink; cut-off = 0.9. Id-Score is the linear discriminant score.

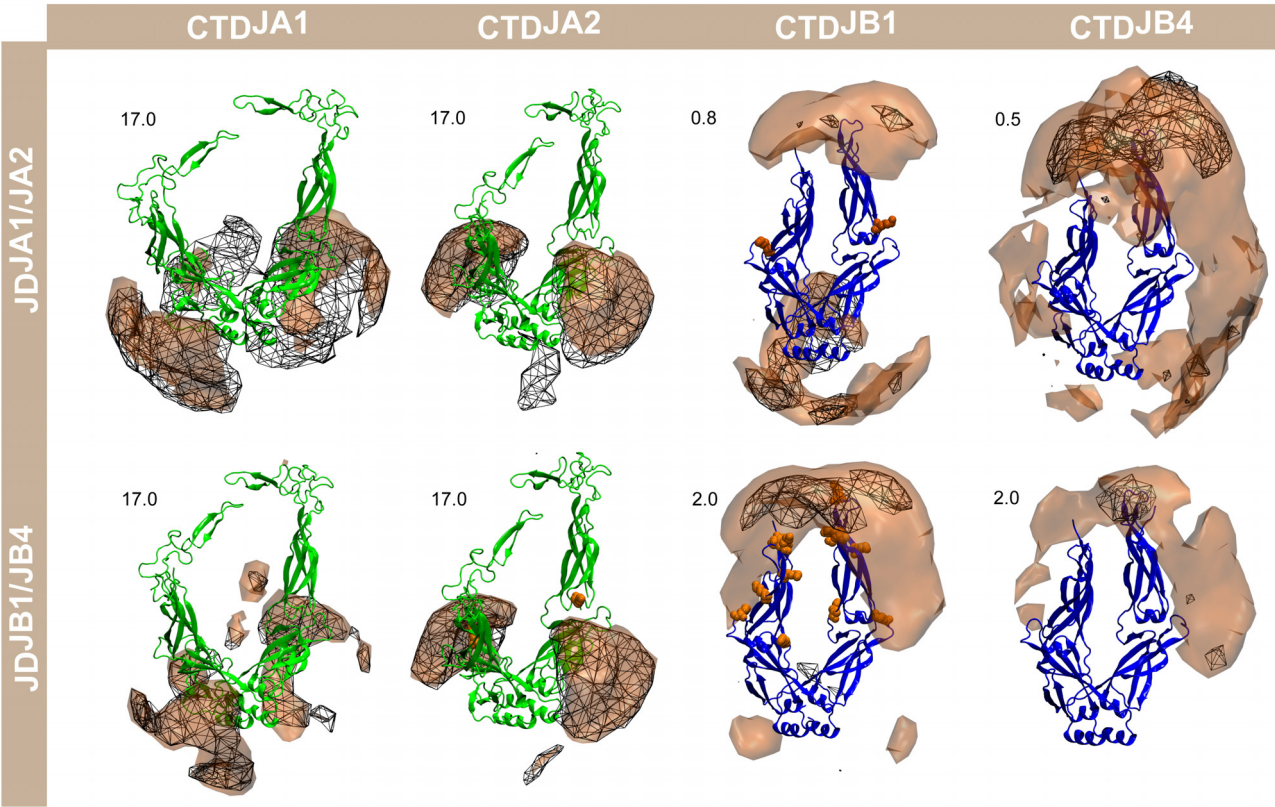
b, Representative mass spectrometry spectra for inter-molecular JA2 and JB1 cross-links. Common peaks, green; cross-linked, red; matched peaks, diamonds (no peaks above 1,100 m/z detected). **c**, SEC profiles of ^3H -labelled JA2 dimer (green cartoon) and ^3H -labelled JB1 dimer (blue cartoon) mixed with unlabelled J-protein from the other class. Precise concentrations are shown in Extended Data Table 1.



Extended Data Figure 6 | Electrostatic interactions between J-domain and CTD predominate in JA2 and JB1 complexes. **a**, FRET efficiencies for JD-CTD and CTD-CTD interactions with 0–0.2% Tween20 titration. Percentage efficiency is relative to untreated (0% Tween20) samples. Donor quenching (black); acceptor fluorescence (red); below, fluorophore positions in J-protein protomers (JA2, green; JB1, blue). N-termini of JD^{JA2} and JD^{JB1} labelled with acceptor fluorophore ReAsH. CTD^{JA2} and CTD^{JB1} labelled with donor fluorophores FlAsH and Alexa Fluor 488 at residues 241 and 278, respectively. **b**, Disaggregation/refolding of preformed luciferase by JA2 and/or JB1 with increasing amounts of Tween20 ($n = 2$). **c**, FRET efficiencies for

JA2 and JB1 interactions at increasing salt concentrations. **d**, Disaggregation/refolding of preformed luciferase aggregates by JA2 and JB1 with increasing salt concentrations; control, 50 mM salt, no chaperones ($n = 2$). **e**, Luciferase disaggregation/refolding in the presence of excess J-domain fragments carrying JD-QPN^{JA2/JB1} mutation of the HPD motif ($n = 3$). **f**, Competition with unlabelled full-length wild-type J-protein (FL); unlabelled competitor is 1–10× acceptor; (-), no competitor. **g**, Competition with unlabelled isolated JD^{JA2} and JD^{JB1}. Data are mean \pm s.e.m., average of at least two experiments for FRET experiments. Precise concentrations are shown in Extended Data Table 1.

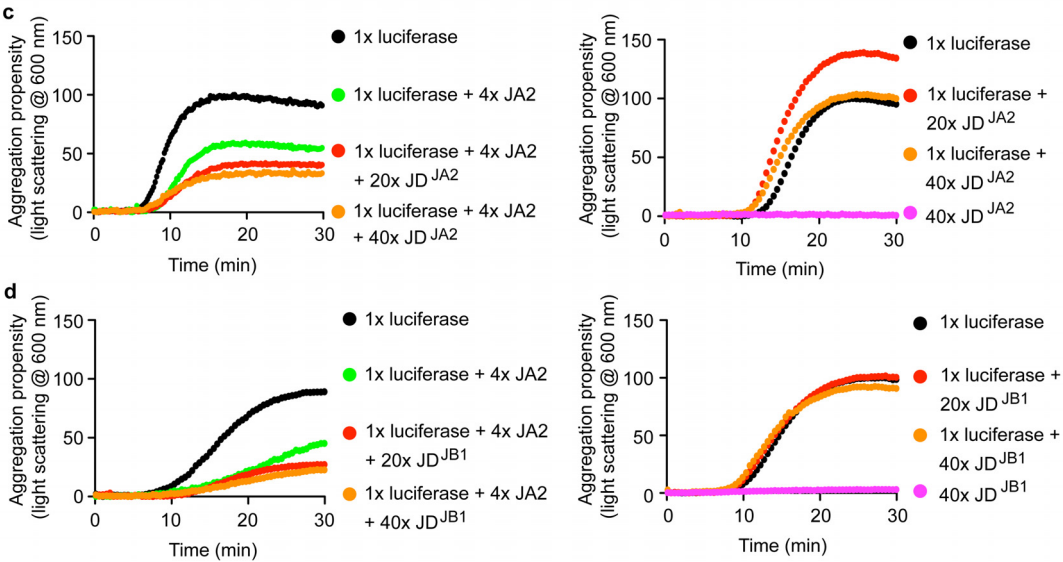
a



b

	CTDJA1	CTDJA2	CTDJB1	CTDJB4
JDJA1	11/25, (-23.6 to -20.8)	6/11, (-25.9 to -24.2)	42/108, (-14.6 to -12.4)	40/109, (-14.9 to -11.5)
JDJA2	4/4, (-23.9 to -22.4)	4/6, (-26.3 to -25.7)	26/65, (-16.0 to -13.9) 94/206, (-11.9 to -7.1)*	23/58, (-17.2 to -13.5)
JDJB1	4/9, (-22.9 to -20.8)	2/4, (-24.3 to -24.0)	21/60, (-16.8 to -14.4)	19/60, (-17.3 to -14.2)
JDJB4	3/9, (-22.1 to -21.5)	6/11, (-25.4 to -23.2)	11/34, (-18.0 to -15.7)	1/8, (-24.5)

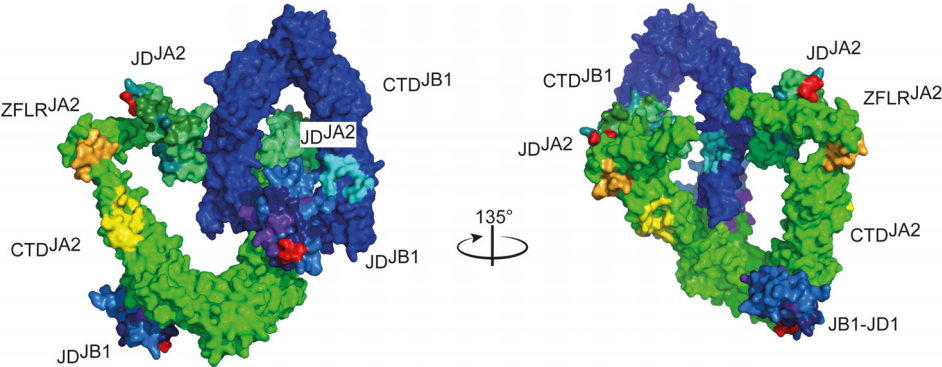
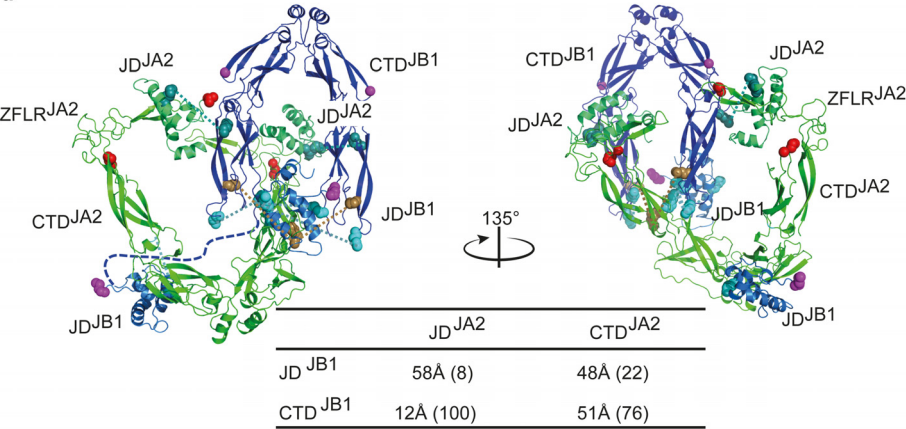
* represents the values obtained for the selected docked complexes that fall within the cross-linking range (30 Å) between Lys209 in CTD of JB1 and Lys46 in JD of JA2



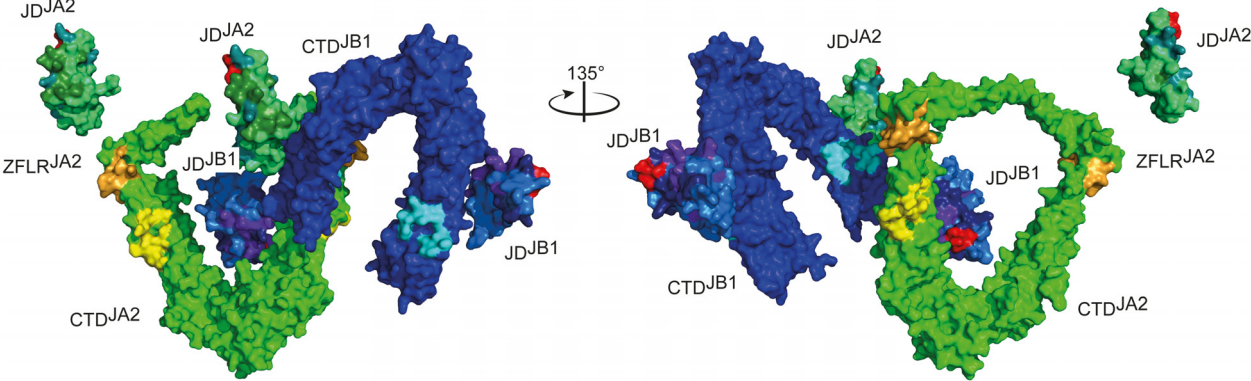
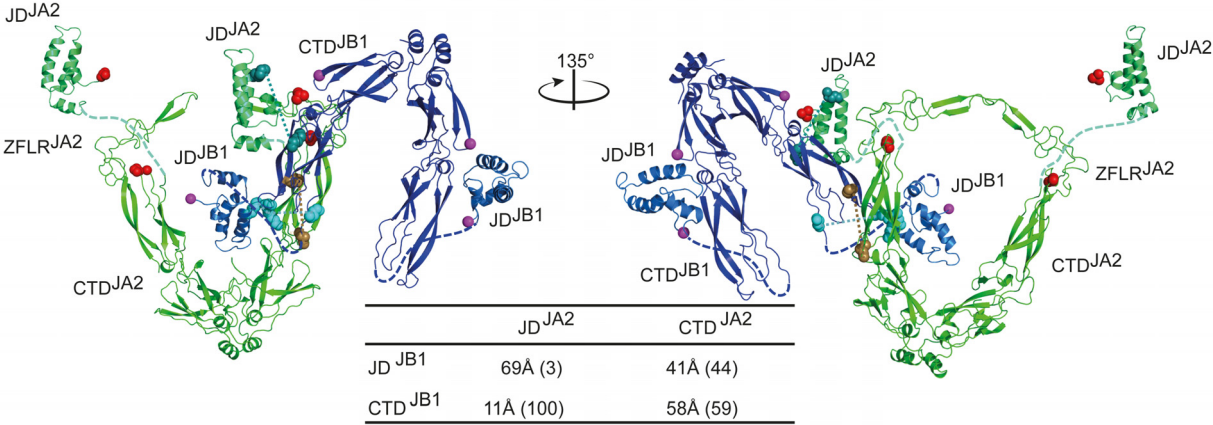
Extended Data Figure 7 | *In silico* prediction of JD–CTD interactions between class A and B J-proteins and *in vitro* evidence that physical interactions between J-proteins do not overlap J-protein substrate binding sites. **a**, Preferred positions of the centres of geometry (CoG) of J-domains (y axis, JA1, JA2, JB1 and JB4) around CTD dimers (x axis, class A, green, class B, blue) obtained from molecular docking simulations. $\text{JD}^{\text{JA1/JB1}}$, wireframe meshes; $\text{JD}^{\text{JA2/JB4}}$, brown contours, each contoured at the isovalue given in the top left of each image. The higher scores for class A CTDs indicate greater specificity of the complexes formed with J-domains; the lower scores for class B CTDs indicate much less specific interactions. Lysines in inter- and intra-J-protein JA2–JB1 cross-links, orange spheres. **b**, Properties of the docking arrangements obtained after clustering. Total number of clusters per simulation, denominator; number of selected clusters (corresponding to 90% of

all docked complexes), numerator, bold. In parentheses, the range of average energy values (in units of kT) for the selected clusters. Lower energy values indicate more favourable binding; fewer clusters indicate a more defined binding mode (see Methods). JD^{JA2} docking to CTD^{JB1} is much weaker and less specific than JD^{JB1} docking to CTD^{JA1} , but docking arrangements compatible with cross-linking results still obtain (Fig. 2d). **c**, Competition of isolated JD^{JA2} fragments against JA2 holdase function in luciferase aggregation at 42 °C ($n = 2$). **d**, Competition of isolated JD^{JB1} fragments against JA2 holdase function ($n = 2$). Luciferase, 1×; JA2, 4×; isolated J-domain fragments, 20× (red; 5-fold excess over JA2), or 40× (orange; 10-fold excess over JA2). Light scattering measured at 600 nm. Precise concentrations are shown in Extended Data Table 1.

a



b



Extended Data Figure 8 | Possible configurations of the JA2–JB1 mixed-class complex. a, Compact configuration. b, Open configuration.

Configurations were derived from computational docking, using constraints from experimental FRET and cross-linking data (Fig. 2a–d and Extended Data Fig. 5a). Each configuration is shown from two views (left and right) rotated by 135 degrees with respect to each other and in ribbon (top) and molecular surface (bottom) representations. In both cases J-domains of JA2 dock onto the CTD dimer of JB1, and similarly J-domains of JB1 dock to the CTD dimer of JA2. Both CTD^{JB1} protomers are within cross-linking distance of CTD^{JA2}. Unstructured glycine/phenylalanine (G/F)-rich flexible regions connecting J-domains and CTDs shown by dark blue (JB1) or green (JA2) dashed lines. Residues at FRET fluorophore sites are shown in space-filling representation (red on JA2, magenta on JB1). Inter-molecular crosslinking lysine pairs (gold and cyan, space-filling) are connected by dotted lines. Bottom left within

a: molecular surface representation of compact configuration of the JA2–JB1 complex, showing substrate binding sites from crystallographic²⁴ (yellow) and biochemical⁵⁷ (orange, cyan) data. HPD motif, red. Residues implicated in JD–HSP70 interactions^{19,58,59} (dark teal and dark green on JD^{JA2}; purple and dark blue on JD^{JB1}). Bottom right within **a**: rotated image. Table shows fluorophore separation distances; calculated percentage FRET efficiencies in parentheses. **a**, Both CTD^{JB1} protomers are within cross-linking distance of CTD^{JA2}. **b**, As in **a**, but with only a single CTD^{JB1} protomer within cross-linking distance to CTD^{JA2}; one JD^{JA2} docks onto CTD^{JB1}, the other JD^{JA2} is free. Similarly, one JD^{JB1} docks onto CTD^{JA2}, the other JD^{JB1} docks onto its own CTD, consistent with SAXS-determination of class B J-proteins^{16,17}. Model of JB1 (blue) based on the crystal structure of CTD and NMR structure of J-domain. Homology model of JA2 (green) based on the crystal structure of Ydj1 (see Methods).

Extended Data Table 1 | Tabulated reaction conditions

Panel	Heat treatment time/ Temp	Luciferase/ Hsp26 rxn concentrations (aggregating or denaturing conditions)	Chaperone mixture	Reaction time/ Temp
Fig.1b	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [2 µM], NEF [0.1 µM], J-protein [1 µM], 0.5 J-protein [0.5 µM]	120 min/ 30°C
Fig.1c	15 min/ 45°C	20 nM (25 nM)	HSP-1 [8 µM], NEF [4 µM], J-protein [4 µM], 0.5 J-protein [2 µM]	120 min/ 20°C
Fig.1d	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], GroEL ^{D87K} [10 µM], J-protein [380 nM], 0.5 J-protein [190 nM]	40 min/ 30°C
Fig.1e,f	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 0.5 J-protein [190 nM]	120 min/ 30°C
Fig.1g	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 0.5 J-protein [190 nM]	40 min/ 30°C
Fig.2e	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 0.5 J-protein [190 nM], 1x JD [380 nM], 5x JD [1.9 µM], 10x JD [3.8 µM]	120 min/ 30°C
Fig.2f	10 min/ 42°C	20 nM/ 100 nM (20 nM/ 100 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 0.5 J-protein [190 nM], 1x JD [380 nM], 5x JD [1.9 µM], 10x JD [3.8 µM]	80 min/ 30°C
Fig.3e	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 0.5 J-protein [190 nM]	120 min/ 30°C
Fig.3f	10 min/ 42°C	20 nM/ 100 nM (20 nM/ 100 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 0.5 J-protein [190 nM]	80 min/ 30°C
Panel	Heat treatment time/ Temp	Luciferase/ Hsp26 rxn concentrations (aggregating or denaturing conditions)	Chaperone mixture	Reaction time/ Temp
EDFig.1c	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	NA	0 min
EDFig.1d	10 min/ 42°C	20 nM/ 100 nM (20 nM/ 100 nM)	HSP70 [750 nM], NEF [40 nM], 0.5 JA2 [190 nM], 0.5 JB1 [190 nM]	0 min
EDFig.1f	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 1/2x J-protein [190 nM], 3x J-protein [1140 nM], 6x J-protein [2280 nM], 9x J-protein [3420 nM]	120 min/ 30°C
EDFig.2a	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 0.5 J-protein [190 nM]	120 min/ 30°C
EDFig.2b	10 min/ 42°C	20 nM/ 100 nM (20 nM/ 100 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 0.5 J-protein [190 nM]	80 min/ 30°C
EDFig.2c	15 min/ 45°C	20 nM (25 nM)	HSP-1 [400 nM], NEF [20 nM], J-protein [200 nM], 0.5 J-protein [100 nM]	120 min/ 20°C
EDFig.2d	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [3 µM], NEF [0.1 µM], J-protein [1 µM], 0.5 J-protein [0.5 µM]	120 min/ 22°C
EDFig.2 e,f,g	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], 100% J-protein [380 nM]	120 min/ 30°C
EDFig.3a	15 min/ 45°C	20 nM (25 nM)	HSP70 [750 nM], NEF [40nM], J-protein [380 nM], 0.5 J-protein [190 nM]	120 min/ 30°C
EDFig.3b	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], J-protein [380 nM], 0.5 J-protein [190 nM]	120 min/ 30°C
EDFig.3e	15 min/ 45°C	20 nM/ 100 nM (2 µM/ 10 µM)	HSP70 [2 µM], NEF [0.1 µM], J-protein [1 µM], 0.5 J-protein [0.5 µM]	120 min/ 30°C
EDFig.3f	15 min/ 45°C	100 nM/ 500 nM (100 nM/ 500 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 0.5 J-protein [190 nM]	120 min/ 30°C
EDFig.3i	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 0.5 J-protein [190 nM]	120 min/ 30°C
EDFig.4b	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], GroEL ^{D87K} [10 µM], J-protein [380 nM], 0.5 J-protein [190 nM]	40 min/ 30°C
EDFig.4e	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], 0.5 J-protein [190 nM]	240 min/ 4-30°C
EDFig.4f	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], J-protein [380 nM], 0.5 J-protein [190 nM], 0.2 J-protein [76 nM], 0.8 J-protein [304 nM]	40-120 min/ 30°C
EDFig.5c	NA	NA	J-protein [190 nM]	10 min/ 30°C
EDFig.6b	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], 0.5 J-protein [190 nM]	120 min/ 30°C
EDFig.6d	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], 0.5 J-protein [190 nM]	90 min/ 30°C
EDFig.6e	15 min/ 45°C	20 nM/ 100 nM (25 nM/ 125 nM)	HSP70 [750 nM], NEF [40 nM], 0.5 J-protein [190 nM], 1x JD [380 nM], 5x JD [1.9 µM], 10x JD [3.8 µM]	120 min/ 30°C
Panel	Heat treatment time/ Temp	α-glucosidase/ Hsp26 rxn concentrations (aggregating conditions)	Chaperone mixture	Reaction time/ Temp
EDFig.3c	15 min/ 50°C	40 nM/ 400 nM (50 nM/ 500 nM)	HSP70 [3 µM], NEF [0.1 µM], J-protein [1 µM], 0.5 J-protein [0.5 µM]	300 min/ 30°C
EDFig.4c,d	15 min/ 50°C	40 nM/ 400 nM (50 nM/ 500 nM)	HSP70 [3 µM], NEF [0.1 µM], J-protein [1 µM], 0.5 J-protein [0.5 µM]	60 min/ 30°C
Panel	Heat treatment time/ Temp	malate dehydrogenase/ Hsp26 rxn concentrations (aggregating conditions)	Chaperone mixture	Reaction time/ Temp
EDFig.3d	30 min/ 47°C	150 nM/ 750 nM (150 nM/ 750 nM)	HSP70 [2 µM], NEF [0.1 µM], 1 J-protein [0.063 µM], 2 J-protein [0.126 µM], 4 J-protein [0.252 µM], GroEL [1 µM], GroES [1 µM]	220 min/ 30°C
Panel	Light scattering	Luciferase	Chaperone mixture	Reaction time/ Temp
EDFig.3g,h	600 nm	200 nM	4x J-protein [800 nM], 4x BSA (control) [800 nM]	30 min/ 42°C
EDFig.7c,d	600 nm	200 nM	4x J-protein [800 nM], 20x JD [4 µM], 40x JD [8 µM]	30 min/ 42°C

Precise concentrations and conditions tabulated for reactions in Figures as indicated in first column (Panel). EDFig., Extended Data Figure.

X-ray structure of a mammalian stearyl-CoA desaturase

Yonghong Bai¹, Jason G. McCoy¹, Elena J. Levin¹, Pablo Sobrado², Kanagalaghatta R. Rajashankar³, Brian G. Fox² & Ming Zhou¹

Stearyl-CoA desaturase (SCD) is conserved in all eukaryotes and introduces the first double bond into saturated fatty acyl-CoAs^{1–4}. Because the monounsaturated products of SCD are key precursors of membrane phospholipids, cholesterol esters and triglycerides, SCD is pivotal in fatty acid metabolism. Humans have two SCD homologues (SCD1 and SCD5), while mice have four (SCD1–SCD4). SCD1-deficient mice do not become obese or diabetic when fed a high-fat diet because of improved lipid metabolic profiles and insulin sensitivity^{5,6}. Thus, SCD1 is a pharmacological target in the treatment of obesity, diabetes and other metabolic diseases⁷. SCD1 is an integral membrane protein located in the endoplasmic reticulum, and catalyses the formation of a *cis*-double bond between the ninth and tenth carbons of stearyl- or palmitoyl-CoA^{8,9}. The reaction requires molecular oxygen, which is activated by a di-iron centre, and cytochrome *b*₅, which regenerates the di-iron centre¹⁰. To understand better the structural basis of these characteristics of SCD function, here we crystallize and solve the structure of mouse SCD1 bound to stearyl-CoA at 2.6 Å resolution. The structure shows a novel fold comprising four transmembrane helices capped by a cytosolic domain, and a plausible pathway for lateral substrate access and product egress. The acyl chain of the bound stearyl-CoA is enclosed in a tunnel buried in the cytosolic domain, and the geometry of the tunnel and the conformation of the bound acyl chain provide a structural basis for the regioselectivity and stereospecificity of the desaturation reaction. The dimetal centre is coordinated by a unique spacial arrangement of nine conserved histidine residues that implies a potentially novel mechanism for oxygen activation. The structure also illustrates a possible route for electron transfer from cytochrome *b*₅ to the di-iron centre.

A Δ2–23 amino-terminal truncation of mouse SCD1 was crystallized in lipidic cubic phase¹¹ and the structure was solved by single-wavelength anomalous dispersion (Extended Data Table 1). SCD1 has four transmembrane helices (TM1–TM4) arranged in a cone-like shape with TM4 sandwiched between TM1 and TM2 (Fig. 1a). Residues in the membrane-spanning region are largely hydrophobic, with the notable exception of a conserved arginine (Arg249; Extended Data Fig. 1) located on TM4 in the centre of the cone (Extended Data Fig. 2). Previous biochemical analysis¹² determined that the amino and carboxy termini are on the cytosolic side of the membrane (Fig. 1b). On the cytosolic side, TM2 and TM4 protrude three helical turns out of the membrane and provide some of the coordinating residues for the dimetal active site. The cytosolic domain comprises 93 residues between TM2 and TM3 (C1) and the 90-residue C terminus (C2) (Fig. 2a). The C1 and C2 domains contain six and five α-helices, respectively. Three of the α-helices (AH1 on C1 and AH7 and AH9 on C2) are amphipathic and probably reside at the interface between the cytosolic domain and the lipid bilayer (Fig. 1a and Extended Data Fig. 3). These amphipathic helices and the locations of hydrophobic residues on the transmembrane helices indicate the approximate position of the lipid bilayer (Fig. 1a). In the crystal lattice, the interacting

surfaces between neighbouring molecules are small (Extended Data Fig. 4), and size-exclusion chromatography (SEC) of the detergent solubilized protein indicated that mouse SCD1 was stable as a monomer. By contrast, previous *in vivo* studies have shown that SCDs are dimers in the cellular membrane¹³. Whether this difference is a consequence of isolation of the enzyme remains to be determined.

The cytosolic domain contains a substantial non-protein density consistent with an 18-carbon acyl-CoA molecule (Fig. 2a and Extended Data Fig. 5a). We modelled a stearyl-CoA molecule into this density, although we were unable to distinguish between oleoyl-CoA and stearyl-CoA solely from the crystallographic maps. The electron density for the CoA moiety is well resolved, and the CoA group interacts primarily with hydrophilic and charged residues on the outer surface of the C1 domain (Fig. 2b). The residues that form polar interactions with the CoA group in the mouse SCD1 structure are strongly conserved among known stearyl-CoA desaturases, including human SCD1, but not among stearyl-lipid desaturases (Extended Data Fig. 1).

The acyl chain is enclosed in a long, narrow tunnel extending approximately 24 Å into the mostly hydrophobic interior of the protein. This tunnel is sharply kinked where it binds to C9 and C10 on stearyl-CoA, the atoms involved in formation of the *cis*-double bond (Fig. 2c). The positioning of C9 and C10 by the kink is enforced by the shape complementarity of this substrate tunnel, by the location of the CoA binding site, and by a hydrogen bond between the Trp258 side chain and the acyl carbonyl (Fig. 2c). The kink in the tunnel is created by the side chains of two conserved residues, Trp149 and Thr257, which are stabilized by hydrogen bonds with conserved Gln143 (Fig. 2c). We note that the narrow and kinked tunnel precisely positions the acyl chain for Δ9-regioselective desaturation, an idea that was envisioned by Bloch over four decades ago⁸.

Previous studies showed that rat SCD1 was effective on acyl chains containing between 14 and 19 carbons, and had the highest activity with substrates 17 to 19 carbons in length⁹. Regardless of the acyl chain length, the double bond is exclusively placed between C9 and C10, as enforced by the substrate–enzyme interactions described above. Residues that probably have a role in determining substrate length are found at the end of the substrate tunnel, which is capped by Tyr104 on TM2 (Fig. 2c). There is a distance of 4.1 Å between the end of the 18:0 acyl chain and the tyrosine hydroxyl oxygen, which agrees well with the observed preference of the enzyme for 18:0 (ref. 9). Tyr104 is highly conserved in animal SCD1. However, an atypical acyl-CoA desaturase (ChDes1) from the marine copepod *Calanus hyperboreus* has a threonine at the position corresponding to Tyr104 in mouse SCD1 (ref. 14). ChDes1 preferentially acts on very long-chain fatty acyl-CoAs (22:0–26:0), but when this threonine was mutated to tyrosine, desaturation of 26:0 was lost while desaturation of 18:0 was retained¹⁴. Another conserved residue, Ala108, is located one helical turn above Tyr104 facing the substrate tunnel (Fig. 2c). Desat2 from *Drosophila melanogaster* has a methionine at this position, and can only accept acyl substrates up to 14 carbons long¹⁵. Combined, these

¹Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. ²Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA. ³NE-CAT and Department of Chemistry and Chemical Biology, Cornell University, Argonne National Laboratory, Argonne, Illinois 60439, USA.

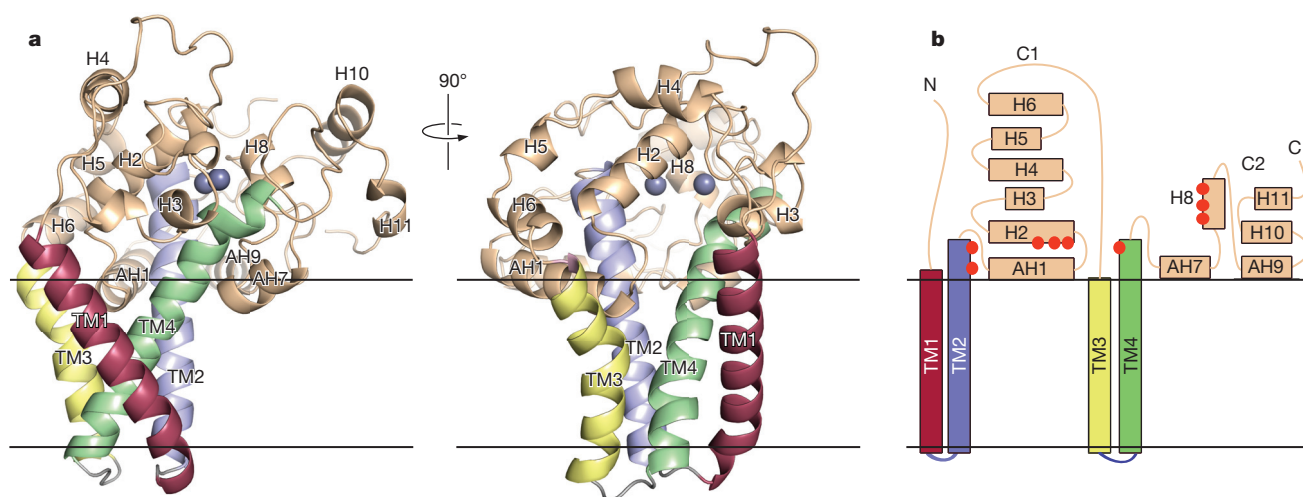


Figure 1 | Structure and topology of mouse SCD1. **a**, The crystal structure of mouse SCD1 is shown from two perpendicular orientations in the membrane, with the cytosolic side on top. Two zinc ions bound to the cytosolic

domain are shown as grey spheres. **b**, Topology diagram of SCD1, with helices coloured by the same scheme as in **a**. Orange spheres represent conserved histidine residues involved in coordination of the dimetal centre.

observations suggest that the tunnel-facing residues 104 and 108 on TM2 are critical determinants of the substrate chain length.

To explore the relationship between the structure of the substrate tunnel in mouse SCD1 and acyl chain selectivity further, we transformed yeast monounsaturated fatty acid auxotroph L8–14C with

either mouse SCD1 or SCD3, which allowed growth in media lacking unsaturated fatty acids. Although SCD1 and SCD3 share 89% primary sequence identity, they yield remarkably different total fatty acid profiles in the yeast host cells, probably reflecting differences in their preferences for reaction with 16:0 and 18:0 (Fig. 2e and ref. 16). In

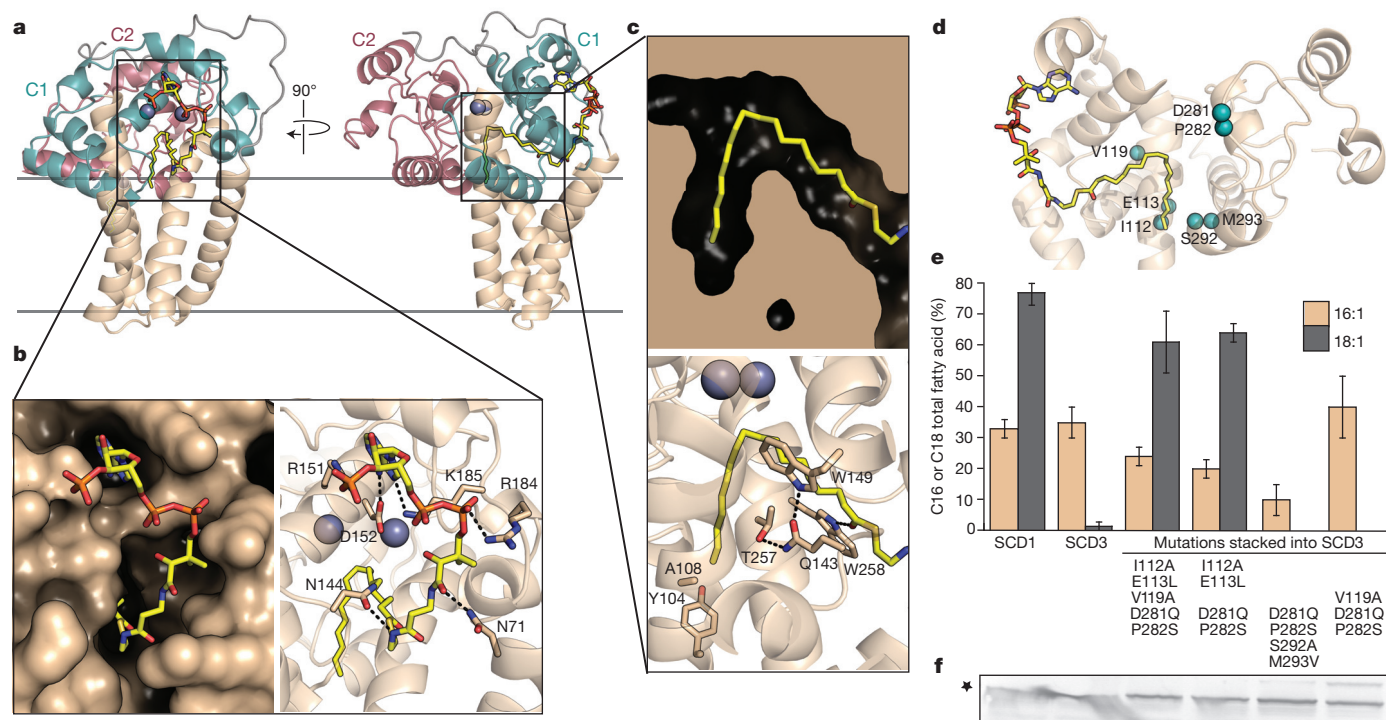


Figure 2 | Architecture of the acyl-CoA binding site. **a**, Two views of the SCD1 structure with the C1 domain coloured teal and the C2 domain coloured magenta. The bound stearoyl-CoA is shown as yellow, red and blue sticks, and the two zinc ions as grey spheres. **b**, A close-up view of the CoA binding site shown as a surface representation (left), or with residues interacting with the CoA moiety shown as sticks (right). **c**, A close-up view of the substrate tunnel housing the acyl chain shown as a surface cross-section (top), or with residues forming the kink (Thr257, Gln143 and Trp149), hydrogen bonding to the acyl oxygen (Trp258), and capping the end of the substrate tunnel (Tyr104 and Ala108) shown as sticks (bottom). **d**, The locations of SCD3 mutations studied by yeast complementation experiments are mapped onto the SCD1 structure

as green spheres. Residue positions are labelled according to the SCD3 sequence. **e**, Monounsaturated fatty acids in the total lipid of yeast L8–14C transformed with mouse SCD1, SCD3 or mutated SCD3 enzyme. SCD1 produces a mixture of 16:1 and 18:1 fatty acids, while SCD3 produces 16:1 nearly exclusively. The combination of Ile112Ala and Glu113Leu mutations converts SCD3 into an enzyme that yields a proportion of monounsaturated fatty acids indistinguishable from SCD1. The corresponding residues in SCD1, Ala108 and Leu109, lie at the end of the substrate binding tunnel. Error bars are s.d. of 3 technical replicates. **f**, Expression levels of the SCD3 mutants were similar as detected by western blotting. Uncropped gel is shown in Extended Data Fig. 6.

SCD1, Ala108, Leu109, Ala288 and Val289 line the distal end of the substrate binding channel, Ala115 is near the position of double bond formation, and Gln277 and Ser278 are on the cytoplasmic surface opposite to the CoA binding site. The corresponding residues in SCD3 are Ile112, Glu113, Ser292 and Met293, Val119, and Asp281 and Pro282, respectively (Fig. 2d). The stacked mutations Ile112Ala/Glu113Leu were able to convert SCD3 from exclusively a 16:0 desaturase into a predominantly 18:0 desaturase (Fig. 2e, f and Extended Data Fig. 6). The stacked mutations Val119Ala/Asp281Gln/Pro282Ser, which are located away from the end of the substrate tunnel, caused no change in the reaction specificity.

In addition to the bound stearyl-CoA molecule, SCD1 also contains two metal ions. The metal ions in the current structure were identified as zinc by X-ray fluorescence, and by diffraction data collected at a wavelength near the zinc absorption edge that yielded two prominent anomalous difference peaks in each protein (Extended Data Fig. 5b–e). Incorporation of zinc instead of iron into the protein was probably an artefact of protein overexpression, and zinc remained the predominant metal species even when the growth media and purification solutions were supplemented with iron.

The dimetal cluster sits at the kink in the substrate tunnel adjacent to C9 and C10 on the substrate, where the double bond is introduced. Zinc 1 (M1) is positioned 5.2 Å from C9, while zinc 2 (M2) is 4.7 Å from C10 (Fig. 3a). M1 and M2 are coordinated by four and five histidine residues, respectively, provided by the helices TM2, TM4, H2 and H8 (Fig. 3b and Extended Data Fig. 7a). The coordination of both zinc ions is consistent with an octahedral geometry with one missing ligand. The nine histidines are highly conserved (Extended Data Fig. 1), and eight of them belong to three histidine-containing motifs (two HXXHH motifs and one HX₄H motif in SCD1) that are characteristic of integral membrane desaturases, alkane hydroxylases and xylene monooxygenases^{3,17}. The predominance of histidine ligands in SCD1 is consistent with the assignment of nitrogen-rich ligation of the di-iron centre in alkane hydroxylase from Mössbauer isomer shifts¹⁸. Mutation of any of these eight histidines into an alanine in rat SCD1 led to a nonfunctional enzyme¹⁷. The ninth histidine (His265) is conserved in other SCDs but had not been previously identified. The water molecule coordinating M1 is hydrogen bonded to an asparagine residue, Asn261 on TM4. Interestingly, this asparagine and His265 belong to a NX₃H motif that is symmetrically equivalent to the HX₄H motif interacting with M2 (Fig. 3c). Likewise, the two HXXHH motifs have symmetrical interactions with M1 and M2 (Fig. 3c).

One notable aspect of the structure is that the two metal ions are separated by 6.4 Å (Fig. 3a). This is longer than in any previously solved structures of soluble di-iron enzymes^{19–21}, including the soluble plant acyl-ACP desaturases that catalyse the same reaction as SCD1. In these soluble enzymes, the two iron ions are bridged by a glutamate

residue with bidentate coordination (Extended Data Fig. 7b) that constrains the inter-iron distance to roughly 3–4 Å (refs 19–21), and permits formation of reaction intermediates such as cis-μ-1,2 peroxo^{22,23} and diferry²⁴. In SCD1, there is no carboxylate coordination between the metal ions. This is probably not an artefact caused by Zn²⁺, as the closest glutamate or aspartate residues are over 6 Å away from the metal centres and provide hydrogen-bonding interactions to the metal-bound His residues (Extended Data Fig. 7a). Since Zn²⁺ has an ionic radius of 0.88 Å, it may have served as a reasonable substitute for Fe²⁺ (ionic radius 0.92 Å) in terms of size and charge during heterologous expression. Notably, the bound Zn²⁺, which often have tetrahedral coordination, have octahedral coordination in the SCD1 structure as typical for iron ions. Given the similar B-values for residues around the observed metal sites, the presence of stearyl-CoA bound in an appropriately kinked configuration relative to the metals, and the absence of reasonably positioned carboxylate residues that might serve as bridging ligands, we propose that the unfortunate presence of Zn²⁺ has not significantly altered the structure and that the SCD1 structure indicates a new avenue for activation of O₂ in biological oxidation reactions.

Two known aspects of the desaturase reaction are compatible with this active site. Removal of the pro-R hydrogen atom from C9 proceeds with k_H/k_D of ~6–7 (refs 9, 25), indicating that this step is rate-limiting. The distance from the pro-R C9 hydrogen atom to the water bound to M1 is 3.5 Å, possibly corresponding to the direction through which the desaturation reaction will initiate. Furthermore, no oxygen atom transfer to carbon is anticipated during the desaturation reaction^{26,27}. The enforcement of a long distance between the acyl chain and metals would be consistent with promotion of an electron transfer mechanism in which O atoms are retained on an oxidizing metal centre as electrons and protons are extracted from the bound and sterically configured acyl chain.

Release of a desaturated acyl chain from the active site merits additional consideration. Given the kink and the narrow aperture of the substrate tunnel, it seems unlikely that substrate entrance or product release can occur by simple linear diffusion in and out of the tunnel. However, a break in the hydrogen bond between Gln143 and Thr257 below the kink in the substrate tunnel would create a fenestration into the hydrophobic core of the membrane (Fig. 2c), allowing lateral transfer of substrates and products into and out of the well-formed substrate tunnel. A separation between TM4 and the loop between helices H1 and H2 could break this bond (Extended Data Fig. 8).

In SCD1, the electrons needed for the desaturation reaction are obtained from cytochrome *b*₅ (cytb5), which in turn obtains electrons from NAD(P)H via cytochrome *b*₅ reductase⁹. Although it is known that cytb5 consists of an N-terminal haem-binding domain and a C-terminal membrane anchor domain²⁸, and must be membrane-anchored to

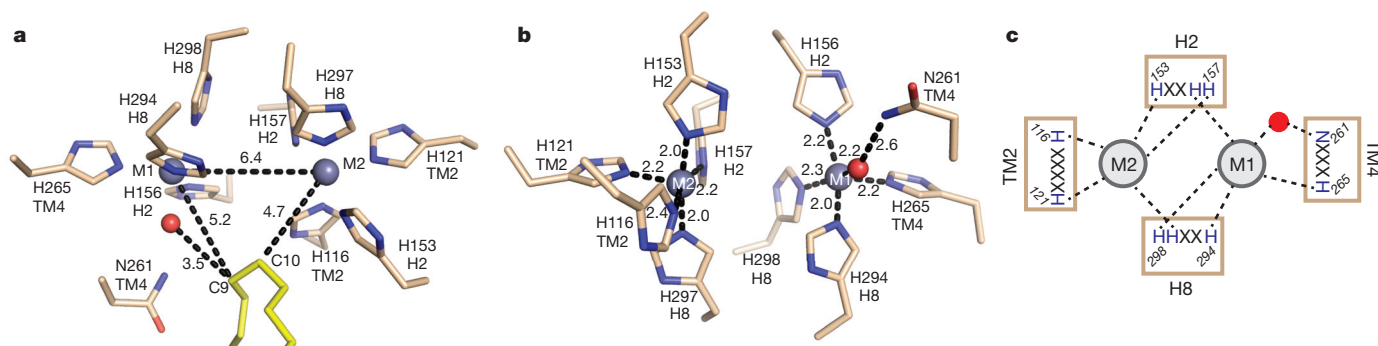


Figure 3 | The dimetal centre. **a, b**, Two views of the dimetal centre and coordinating residues, marked with distances between the zinc ions and C9 and C10 on the substrate (**a**), and coordination distances (**b**). The zinc ions and an ordered water molecule are shown as grey and red spheres, respectively. See

Extended Data Fig. 7 for a stereo view. **c**, Schematic showing the locations of the coordinating His and Asn residues in four conserved motifs on TM2, TM4, H2 and H8.

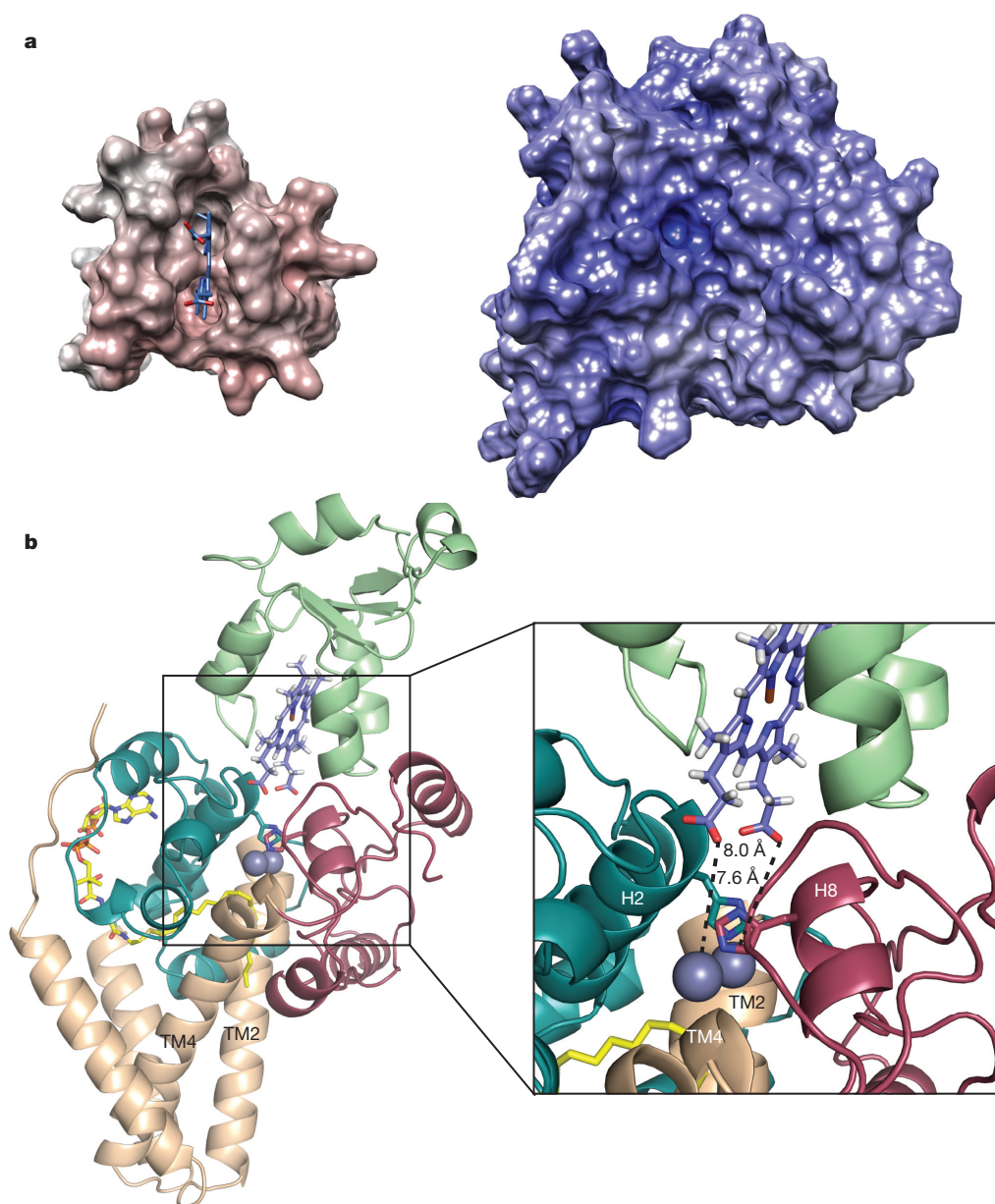


Figure 4 | Proposed interactions between cytb5 and SCD1. **a**, Electrostatic surfaces of cytb5 (left) and mouse SCD1 (right) oriented with their proposed interaction surfaces towards the viewer. **b**, A proposed model of the complex between cytb5 and SCD1. Cytb5 is coloured green; domains C1 and C2 on SCD1 are teal and magenta, respectively. The inset shows a closer view of the proposed electron transfer pathway, with the haem domain shown as blue

function in the desaturation reaction¹, the cytb5 binding site on SCD is unknown. Examination of the mouse SCD1 structure shows that the dimetal centre could be accessible from the cytoplasm via a groove formed between the soluble domains C1 and C2. In the crystal structure, the N terminus of the protein lies along this groove (Extended Data Fig. 9); however, it forms few interactions with the cytoplasmic domain and potentially could be displaced by cytb5. Electrostatic calculations on the two proteins demonstrate that the predominantly positive surface of SCD1 is well complemented by the mostly negative surface of cytb5 (Fig. 4a). Placement of the cytb5 functional domain along this groove would place the electron donor groups within 14 Å of the dimetal centre, an acceptable distance for electron transfer between biological redox centres²⁹ (Fig. 4b). His157 (H2) and His298 (H8) sit directly above the two metal ions and are potential candidates to form an electron transfer interface. This

sticks and the metal ions as grey spheres. The last residue resolved in the cytb5 structure is approximately 35 Å from the predicted position of the bilayer. The 18-residue linker from this residue to the transmembrane helix of cytb5 should therefore be long enough to allow for the haem-containing domain to bind in this orientation.

proposed complex also places negatively charged residues on cytb5 demonstrated to be necessary for complex formation²⁸ close to positively charged residues on H4 of SCD. The structure of mouse SCD1 enables us to address questions related to basic mechanisms of the desaturation reaction more precisely.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 September 2014; accepted 5 May 2015.

Published online 22 June 2015.

- Goren, M. A. & Fox, B. G. Wheat germ cell-free translation, purification, and assembly of a functional human stearyl-CoA desaturase complex. *Protein Expr. Purif.* **62**, 171–178 (2008).
- Paton, C. M. & Ntambi, J. M. Biochemical and physiological function of stearyl-CoA desaturase. *Am. J. Physiol. Endocrinol. Metab.* **297**, E28–E37 (2009).

3. Sperling, P., Ternes, P., Zank, T. K. & Heinz, E. The evolution of desaturases. *Prostaglandins Leukot. Essent. Fatty Acids* **68**, 73–95 (2003).
4. Strittmatter, P. *et al.* Purification and properties of rat liver microsomal stearyl coenzyme A desaturase. *Proc. Natl Acad. Sci. USA* **71**, 4565–4569 (1974).
5. Gutiérrez-Juárez, R. *et al.* Critical role of stearyl-CoA desaturase-1 (SCD1) in the onset of diet-induced hepatic insulin resistance. *J. Clin. Invest.* **116**, 1686–1695 (2006).
6. Ntambi, J. M. *et al.* Loss of stearyl-CoA desaturase-1 function protects mice against adiposity. *Proc. Natl Acad. Sci. USA* **99**, 11482–11486 (2002).
7. Zhang, Z., Dales, N. A. & Winther, M. D. Opportunities and challenges in developing stearyl-coenzyme A desaturase-1 inhibitors as novel therapeutics for human disease. *J. Med. Chem.* **57**, 5039–5056 (2014).
8. Bloch, K. Enzymatic synthesis of monounsaturated fatty acids. *Acc. Chem. Res.* **2**, 193–202 (1969).
9. Enoch, H. G., Catala, A. & Strittmatter, P. Mechanism of rat liver microsomal stearyl-CoA desaturase. Studies of the substrate specificity, enzyme-substrate interactions, and the function of lipid. *J. Biol. Chem.* **251**, 5095–5103 (1976).
10. Behrouzian, B. & Buist, P. H. Fatty acid desaturation: variations on an oxidative theme. *Curr. Opin. Chem. Biol.* **6**, 577–582 (2002).
11. Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P. & Landau, E. M. X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science* **277**, 1676–1681 (1997).
12. Man, W. C., Miyazaki, M., Chu, K. & Ntambi, J. M. Membrane topology of mouse stearyl-CoA desaturase 1. *J. Biol. Chem.* **281**, 1251–1260 (2006).
13. Lou, Y. & Shanklin, J. Evidence that the yeast desaturase Ole1p exists as a dimer *in vivo*. *J. Biol. Chem.* **285**, 19384–19390 (2010).
14. Meesapyodsuk, D. & Qiu, X. Structure determinants for the substrate specificity of acyl-CoA $\Delta 9$ desaturases from a marine copepod. *ACS Chem. Biol.* **9**, 922–934 (2014).
15. Dallerac, R. *et al.* A $\Delta 9$ desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **97**, 9449–9454 (2000).
16. Miyazaki, M., Bruggink, S. M. & Ntambi, J. M. Identification of mouse palmitoyl-coenzyme A $\Delta 9$ -desaturase. *J. Lipid Res.* **47**, 700–704 (2006).
17. Shanklin, J., Whittle, E. & Fox, B. G. Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. *Biochemistry* **33**, 12787–12794 (1994).
18. Shanklin, J., Achim, C., Schmidt, H., Fox, B. G. & Munck, E. Mossbauer studies of alkane omega-hydroxylase: evidence for a diiron cluster in an integral-membrane enzyme. *Proc. Natl Acad. Sci. USA* **94**, 2981–2986 (1997).
19. Högborn, M., Huque, Y., Sjöberg, B. M. & Nordlund, P. Crystal structure of the di-iron/radical protein of ribonucleotide reductase from *Corynebacterium ammoniagenes*. *Biochemistry* **41**, 1381–1389 (2002).
20. Lindqvist, Y., Huang, W., Schneider, G. & Shanklin, J. Crystal structure of delta9 stearyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. *EMBO J.* **15**, 4081–4092 (1996).
21. Sazinsky, M. H. & Lippard, S. J. Correlating structure with function in bacterial multicomponent monooxygenases and related diiron proteins. *Acc. Chem. Res.* **39**, 558–566 (2006).
22. Broadwater, J. A., Ai, J., Loehr, T. M., Sanders-Loehr, J. & Fox, B. G. Peroxidiferic intermediate of stearyl-acyl carrier protein $\Delta 9$ desaturase: oxidase reactivity during single turnover and implications for the mechanism of desaturation. *Biochemistry* **37**, 14664–14671 (1998).
23. Moënné-Loccoz, P., Baldwin, J., Ley, B. A., Loehr, T. M. & Bollinger, J. M. Jr. O_2 activation by non-heme diiron proteins: identification of a symmetric μ -1,2-peroxide in a mutant of ribonucleotide reductase. *Biochemistry* **37**, 14659–14663 (1998).
24. Banerjee, R., Proshlyakov, Y., Lipscomb, J. D. & Proshlyakov, D. A. Structure of the key species in the enzymatic oxidation of methane to methanol. *Nature* **518**, 431–434 (2015).
25. Behrouzian, B. *et al.* Mechanism of fatty acid desaturation in the green alga *Chlorella vulgaris*. *Eur. J. Biochem.* **268**, 3545–3549 (2001).
26. Buist, P. H., Behrouzian, B., Kostas, A. A., Dawson, B. & Black, B. Fluorinated fatty acids: new mechanistic probes for desaturases. *Chem. Commun.* **23**, 2671–2672 (1996).
27. Light, R. J., Lennarz, W. J. & Bloch, K. The metabolism of hydroxystearic acids in yeast. *J. Biol. Chem.* **237**, 1793–1800 (1962).
28. Dailey, H. A. & Strittmatter, P. Characterization of the interaction of amphipathic cytochrome b5 with stearyl coenzyme A desaturase and NADPH:cytochrome P-450 reductase. *J. Biol. Chem.* **255**, 5184–5189 (1980).
29. Page, C. C., Moser, C. C., Chen, X. & Dutton, P. L. Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature* **402**, 47–52 (1999).

Acknowledgements This work was supported by the US National Institutes of Health (R01DK088057, R01GM098878, R01HL086392, U54GM095315, U54GM094584 and R01GM050853), the American Heart Association (12EIA8850017), and the Cancer Prevention and Research Institute of Texas (R12MZ). Final data were collected at Northeastern Collaborative Access Team (NE-CAT) beamlines, which are supported by a grant from the National Institute of General Medical Sciences (P41GM103403). Crystals were screened at beamline 17-ID at the Advanced Photon Source, beamlines 8.2.2 and 5.0.2 at Berkeley Center for Structural Biology at the Lawrence Berkeley Laboratory.

Author Contributions M.Z., B.G.F. and Y.B. conceived the project. Y.B., J.G.M. and E.J.L. expressed, purified and crystallized mouse SCD1, and solved and refined the structure. P.S. and B.G.F. designed and performed mutagenesis and yeast complementation experiments. K.R.R. advised on data collection and structure determination. All authors analysed data and wrote the manuscript.

Author information Atomic coordinates and structure factors have been deposited with the Protein Data Bank under accession number 4YMK. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.Z. (mzhou@bcm.edu) or B.G.F. (bgfox@biochem.wisc.edu).

METHODS

No statistical methods were used to predetermine sample size.

Desaturase expression and purification. To identify an optimal SCD candidate for crystallization, we tested 24 eukaryotic homologues for overexpression in High Five (*Trichoplusia ni*) insect cells using the Bac-to-Bac system (Invitrogen) to generate virus. During the process of identifying suitable homologs, we found that the InterPro fatty acid desaturase family 1 motif³⁰, G-E-X-[FYN]-H-N-[FY]-H-H-X-F-P-X-D-Y, is found in more than 7,000 eukaryotic sequences (fatty acid desaturase type I, InterPro accession IPR005804) from animals, plants and fungi. The three His residues from this motif are now known to provide ligands to both metal ions, and so provide an excellent indicator for the presence of a mouse SCD1-like protein fold. Inclusion of bacterial genes having a related three-His motif expands to more than 22,000 sequences with 90 distinct domain architectures where a mouse SCD1-like fold is fused to domains with a redox centre such as cytb5, plant-type and Rieske-type ferredoxins, rubredoxins and thioredoxins, as well as domains for peptidases, hydrolases, kinases, phosphatases, and many others. This diversity of domain architectures suggests a considerably greater versatility for the SCD1 fold than the originally anticipated desaturase and alkane hydroxylase activities¹⁷.

The pFastBac vector was modified to include a tobacco etch virus (TEV) protease cleavage site before the C-terminal polyhistidine tag. The majority of the 24 homologues expressed, but the mouse SCD1 clone (GI:13938635, Life Technologies Open Biosystems) gave the highest yield and stability. Since full-length mouse SCD1 was prone to aggregation and precipitation, we designed five more mouse SCD1 constructs of various N-terminal truncations based on secondary structure prediction and sequence conservation. The construct containing residues 24–355 was stable and yielded diffracting crystals.

Cells were infected with baculovirus at a density of $\sim 3 \times 10^6$ cells ml⁻¹ and were grown at 27 °C for 48–56 h before being collected by centrifugation at 2,000g for 20 min. Cell membranes were isolated from cell pellets following a published protocol³¹. In brief, cell pellet from 1 l of culture was lysed in 50 ml hypotonic 10 mM HEPES, pH 7.5, containing 10 mM NaCl, 5 mM MgCl₂ and 25 µg ml⁻¹ DNase I. After centrifugation at 55,000g for 45 min, cell membranes were washed with 50 ml high-osmotic buffer containing 25 mM HEPES, pH 7.5 and 1 M NaCl. Purified membranes were re-suspended in a low-osmotic buffer containing 25 mM HEPES, pH 7.5, 150 mM NaCl and 40% (v/v) glycerol, flash frozen with liquid N₂, and stored at -80 °C.

Purified membranes were thawed and dounced in (10 ml per gram membrane) 20 mM HEPES, pH 7.5, 150 mM NaCl and 2 mM β-mercaptoethanol, and solubilized with 2% (w/v) *n*-decyl-β-D-maltoside (Anatrace) at 4 °C for 2 h. After centrifugation (55,000g, 45 min, 4 °C), the desaturase was purified from the supernatant using a cobalt-based affinity resin (Talon, Clontech) and the His-tag was cleaved by TEV protease (leaving an extra ENLYFQ peptide at the C terminus). Purified desaturase was collected and concentrated to 5 mg ml⁻¹ (Amicon 50 kDa cutoff, Millipore) and loaded onto a size-exclusion column (Superdex 200 10/300 GL, GE Health Sciences) equilibrated with 25 mM HEPES, pH 7.5, 150 mM NaCl, 0.18% (w/v) *n*-decyl-β-D-maltoside and 5 mM β-mercaptoethanol. The peak fractions containing desaturase were pooled and immediately used for crystallization. **Lipidic cubic phase crystallization.** Crystallization trials with detergent-solubilized protein and the bicelle method³² failed to yield crystals, whereas the *in meso* method³³ succeeded. For lipidic cubic phase crystallization, the purified desaturase was concentrated to 50 mg ml⁻¹ and two volumes of protein solution and three volumes of molten monoolein (Sigma) were mixed with a coupled syringe device. Crystallization trials were performed with 96-well glass sandwich plates (Molecular Dimensions) and a Gryphon crystallization robot using 50 nl protein-lipid mixture overlaid with 800 nl precipitant solution in each well. The initial crystal hits were systematically optimized by screening against salt and PEG concentrations, pH values, and different lipids. The best crystals grew to a final size of $\sim 50 \times 50 \times 20$ µm within 5 days in 100 mM MES buffer, pH 6.7–7.1, containing 33–37% (v/v) PEG 400, 200 mM NaCl, 4% (v/v) ethylene glycol. Crystals were collected directly from the protein-lipid mixture using 50 µm MiTeGen MicroMounts and immediately flash frozen in liquid nitrogen.

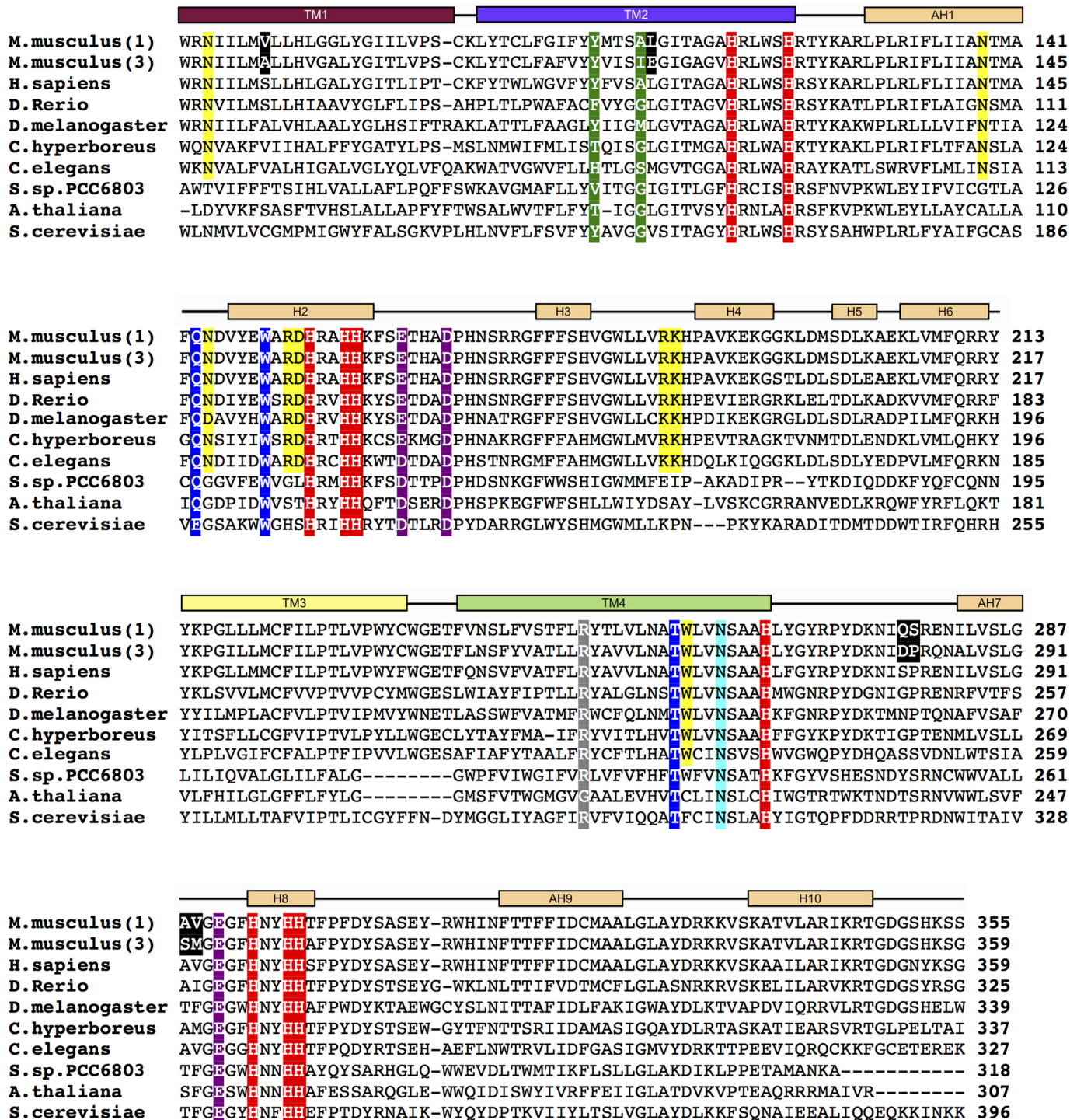
X-ray data collection and processing. X-ray diffraction data were collected at beamline 24ID-C (NE-CAT) at the Advanced Photon Source at Argonne National Laboratory. A data set collected from a single crystal at a wavelength of 1.254 Å with a resolution of 2.8 Å was used for phasing (Extended Data Table 1). The phasing data were processed with XDS³⁴ and scaled with AIMLESS³⁵. The presence of zinc as the predominant ion was confirmed via fluorescence emission spectra using an Amptek SDD fluorescence detector, and analysis of the anomalous signal. A second data set with a resolution of 2.6 Å was collected on a second crystal at a wavelength of 0.9795 Å for use in molecular replacement. The higher resolution

data were indexed, integrated, and scaled using HKL2000 (ref. 36). The crystals belonged to the space group $P2_12_12_1$ with unit cell dimensions of $a = 77.06$ Å, $b = 113.77$ Å and $c = 141.70$ Å.

Structure determination and refinement. SHELXD/SHELXE³⁷ found four anomalous scatterers per asymmetric unit and the resulting density modified map was used to build an all-helical partial model of one protein molecule. The phased translation function from MOLREP³⁸ was then used to locate a second protein molecule. This partial model and the heavy atom sites were input into PHASER-EP³⁹, which was run in MR-SAD phasing mode. Further density modification was carried out with RESOLVE⁴⁰. The structure model was then further built through successive rounds with COOT⁴¹ and refinement with phenix.refine⁴². This model was then used as the input model for molecular replacement using PHASER with the 2.6 Å diffraction data. The model was completed through successive rounds of model building with COOT and refinement with phenix.refine. The crystallographic map was easily interpretable. In the final stages of refinement, TLS groups were determined using TLSMD⁴³ and protein geometry was validated with MolProbity⁴⁴. Figures were produced with PyMOL (Schrödinger LLC.). Electrostatic surfaces were generated with Chimera⁴⁵. For Fig. 4, coordinates were used from rat cytochrome *b*₅ (PDB accession 1BFX; ref. 46), which is 100% identical to mouse cytochrome *b*₅ over the depicted residue range.

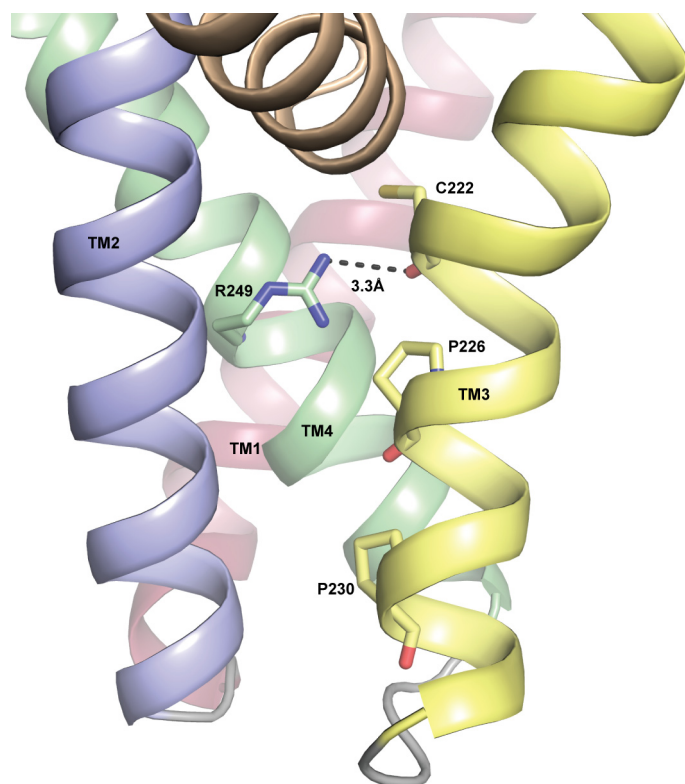
Functional studies of mouse SCDs. For expression using the galactose-inducible yeast expression vector pYES-DEST52 (Invitrogen),⁴⁷ mouse SCD1 and SCD3 genes were cloned with an 81 nucleotide leader sequence encoding the N terminus of *Saccharomyces cerevisiae* desaturase (ole1) appended to the 5' end. Mutations of mouse SCD3 were made using QuikChange (Stratagene). Expression plasmids were transformed into L8-14C (ref. 48), an ole1⁻ yeast strain, and cultured on agar plates containing 0.5 mM each of oleic acid and palmitoleic acid¹⁷. SCD expression was induced on agar plates containing 2% galactose. SCD expression was detected by western blotting using a His-Tag monoclonal antibody (Novagen, 70796-3), goat anti-mouse IgG AP conjugate secondary antibody (Novagen, 69266-3), and alkaline phosphatase reagent (Novagen). Total fatty acids were determined as methyl esters using gas chromatography and mass spectroscopy⁴⁹.

30. Mitchell, A. et al. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43**, D213–D221 (2015).
31. Jaakola, V. P. et al. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **322**, 1211–1217 (2008).
32. Ujwal, R. & Bowie, J. U. Crystallizing membrane proteins using lipidic bicelles. *Methods* **55**, 337–341 (2011).
33. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
34. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
35. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
36. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data. *Methods Enzymol.* **276**, 307–326 (1997).
37. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485 (2010).
38. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010).
39. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
40. Wang, J. W. et al. SAD phasing by combination of direct methods with the SOLVE/RESOLVE procedure. *Acta Crystallogr. D* **60**, 1244–1253 (2004).
41. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
42. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
43. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62**, 439–450 (2006).
44. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
45. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
46. Arnesano, F., Banci, L., Bertini, I., Felli, I. C. & Koulougliotis, D. Solution structure of the B form of oxidized rat microsomal cytochrome *b*₅ and backbone dynamics via 15N rotating-frame NMR-relaxation measurements. Biological implications. *Eur. J. Biochem.* **260**, 347–354 (1999).
47. West, R. W. Jr, Yocum, R. R. & Ptashne, M. *Saccharomyces cerevisiae* GAL1–GAL10 divergent promoter region: location and function of the upstream activating sequence UASG. *Mol. Cell. Biol.* **4**, 2467–2478 (1984).
48. Stuke, J. E., McDonough, V. M. & Martin, C. E. Isolation and characterization of OLE1, a gene affecting fatty acid desaturation from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **264**, 16537–16544 (1989).
49. Gomez, F. E. et al. Molecular differences caused by differentiation of 3T3-L1 preadipocytes in the presence of either dehydroepiandrosterone (DHEA) or 7-oxo-DHEA. *Biochemistry* **41**, 5473–5482 (2002).

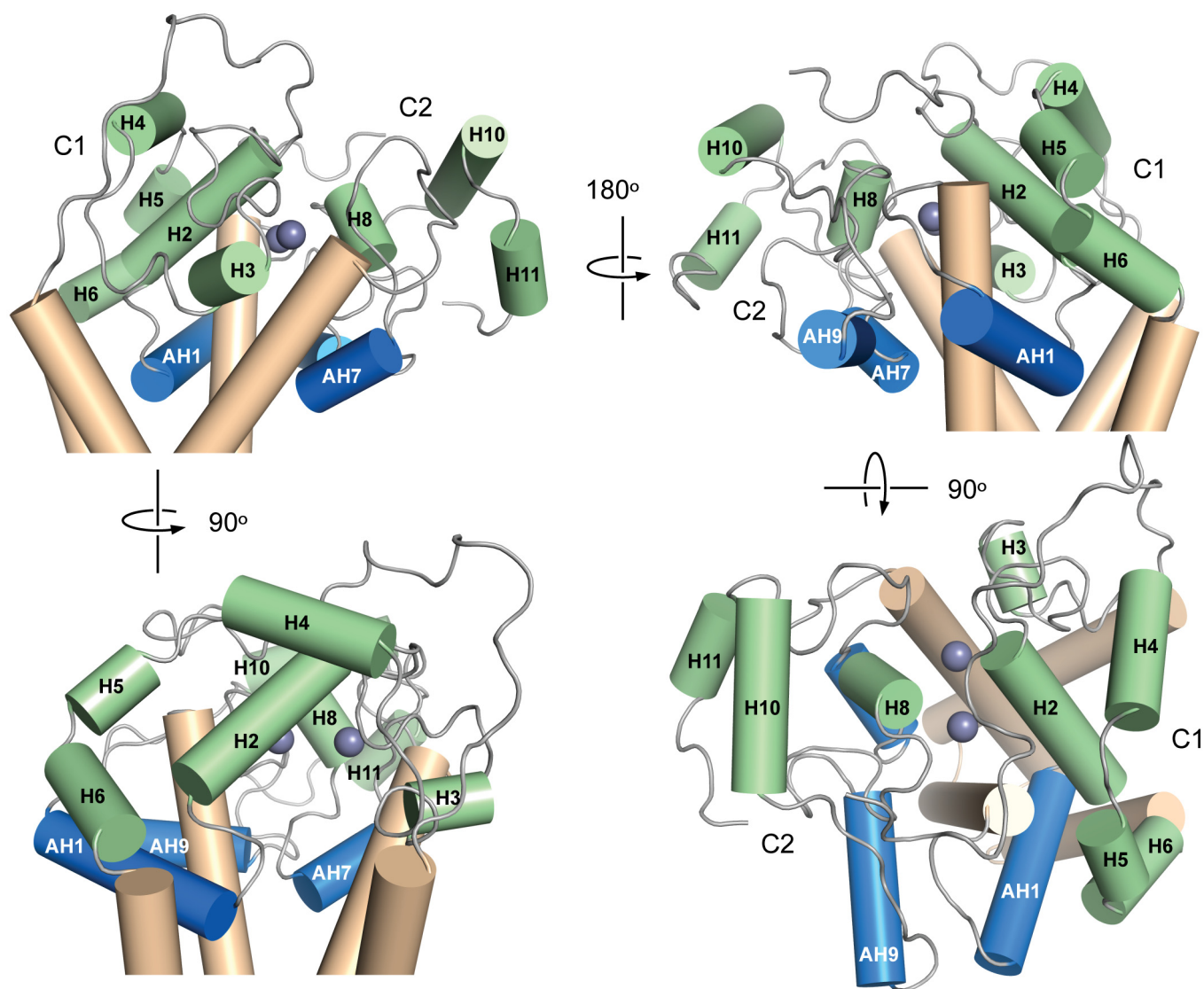


Extended Data Figure 1 | Sequence alignment of mouse SCD1 with other integral membrane desaturases. The N terminus of mouse SCD1 is not shown. For all the other sequences, only the region aligning to mouse SCD1 is included. Secondary structure elements from the mouse SCD1 crystal structure are labelled. Residues discussed in the text are highlighted in red (histidines in the primary coordination sphere of the dimetal unit), purple (carboxylates in the secondary coordination sphere of the dimetal unit), blue (acyl-chain binding site), yellow (CoA-binding site), green (residues that may determine the length of bound acyl chains), black (mutations that change the substrate

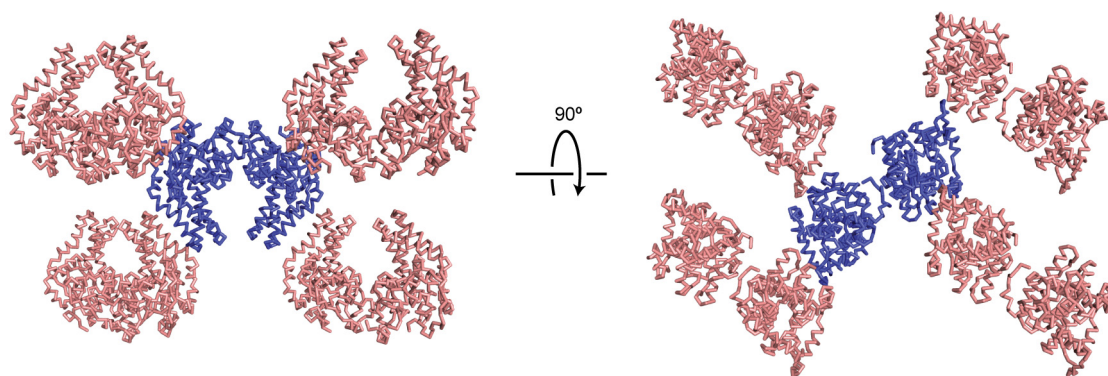
specificity in mouse SCD3) and grey (Arg249 in the transmembrane region). The accession numbers for sequences included in the alignment are: mouse SCD1 (GI: 31543675), mouse SCD3 (GI: 13277368), human SCD1 (GI: 53759151), zebrafish SCD1 (GI: 28394115), *D. melanogaster* desat2 (GI: 24646295), *C. hyperboreus* ChDes9-1 (GI: 589834955), *C. elegans* FAT5 (GI: 544604099), delta-9 desaturase from *Synechocystis* sp. PCC 6803 (GI: 339274799), delta-9 desaturase from *A. thaliana* (GI: 18402641), and yeast OLE1 (GI: 1322552).



Extended Data Figure 2 | Structural role of Arg249. The conserved arginine residue Arg249, located on TM4 within the transmembrane region of the protein, forms a hydrogen bond with the carbonyl oxygen of Cys222 on TM3. This interaction may help stabilize the kink in TM3 caused by Pro226 on the following turn.

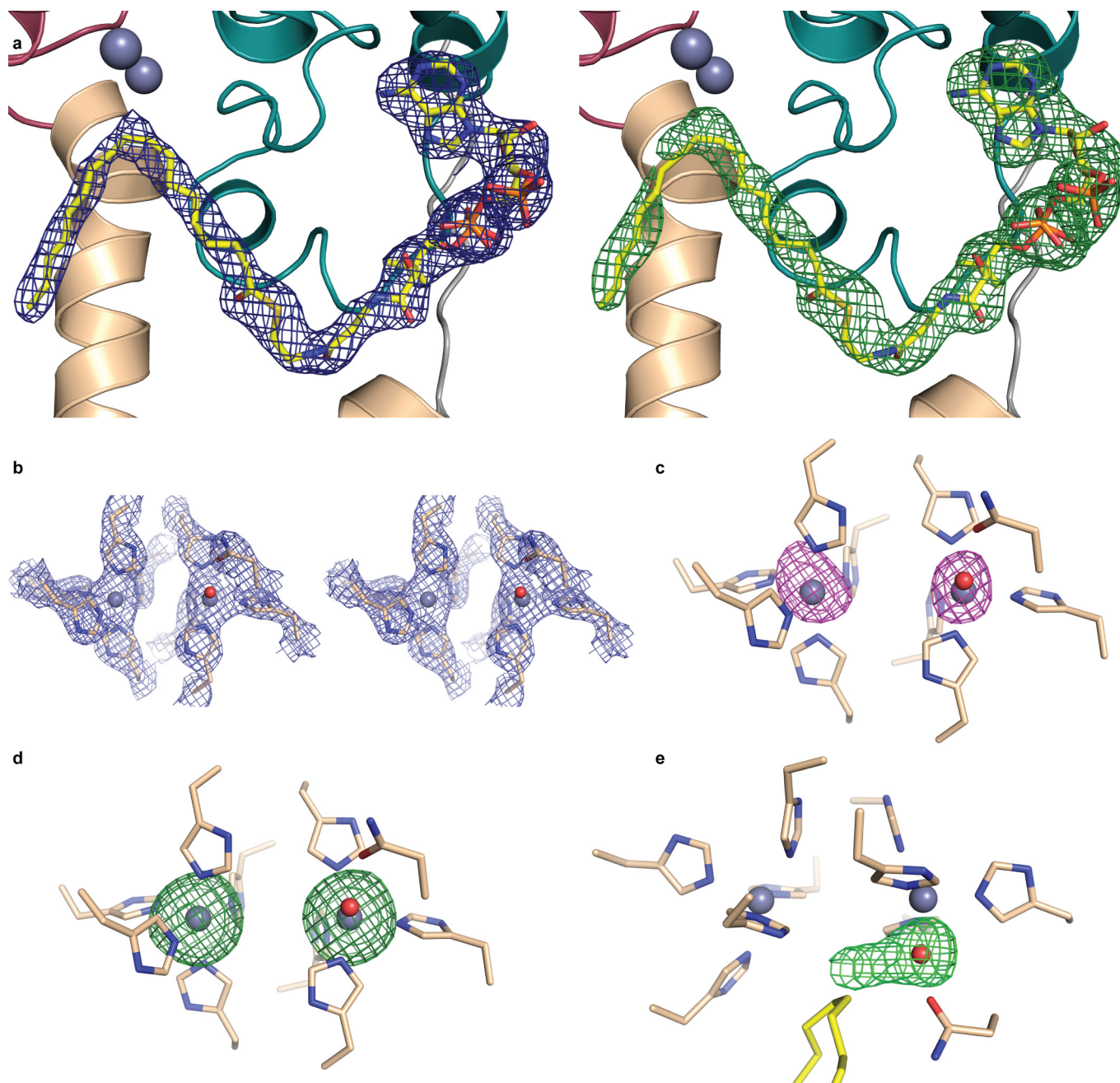


Extended Data Figure 3 | Structure of the SCD1 cytoplasmic domain. Four views of the cytoplasmic domain. The proposed amphipathic helices are coloured blue, while the other helices forming the cytoplasmic domain are green.



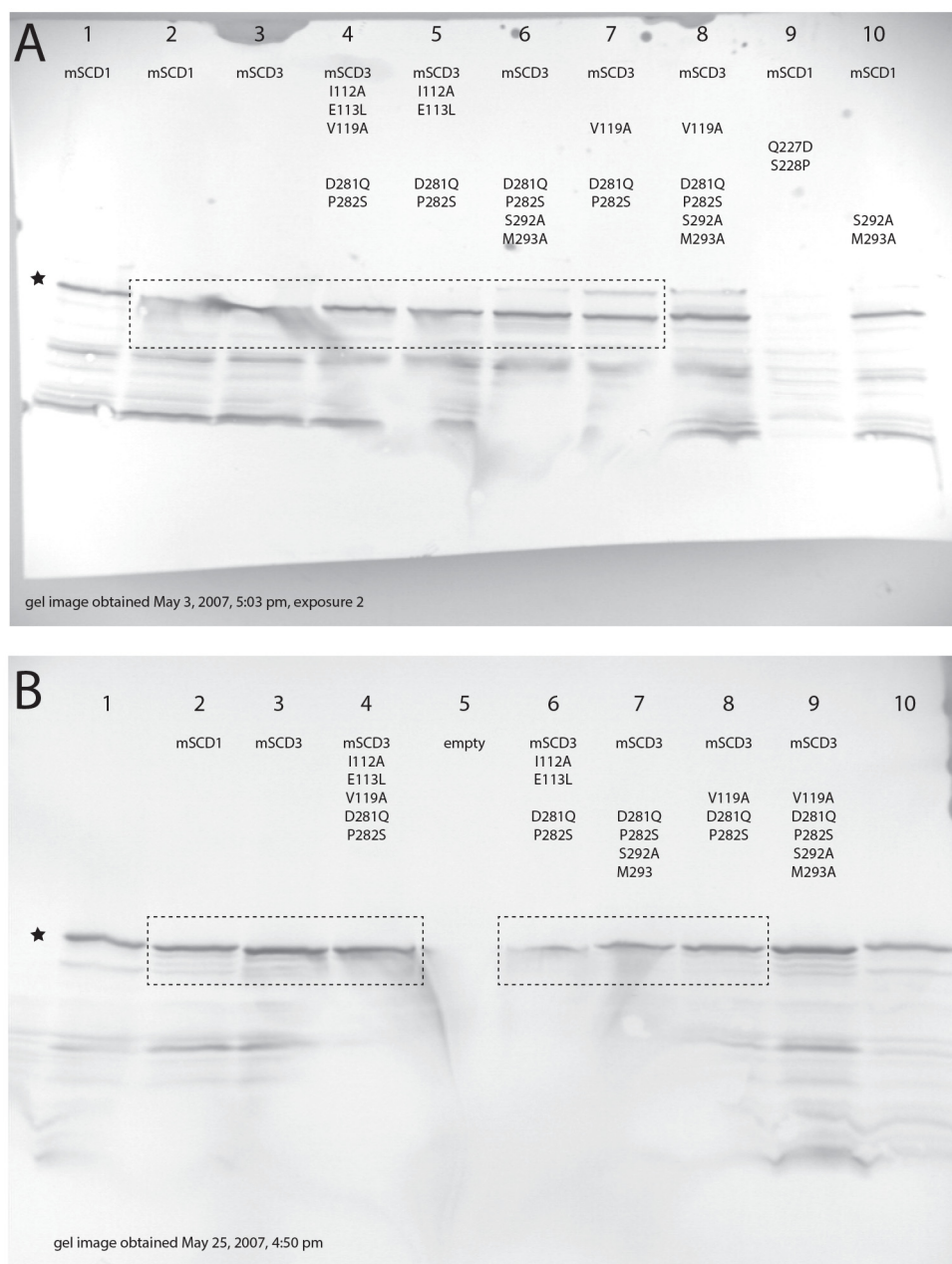
Extended Data Figure 4 | The mouse SCD1 crystal lattice. Cross-sections of the crystal lattice for the $P2_12_12_1$ mouse SCD1 lipidic cubic phase crystals, viewed from two perpendicular directions. One asymmetric unit is coloured blue. Within an individual asymmetric unit, interactions between the two

chains are mediated by residues from a C-terminal cloning artefact. All interactions with chains in neighbouring asymmetric units involve antiparallel orientations of the interacting monomers and have small interface areas.



Extended Data Figure 5 | Electron density maps for acyl-CoA and the dimetal centre. **a**, Stearoyl-CoA bound to SCD1 is superposed with the weighted $2F_o - F_c$ electron density contoured at 1.5σ (left) or $F_o - F_c$ electron density calculated with the substrate molecule omitted and contoured at 2.3σ (right). **b**, Stereoview of the dimetal centre and coordinating histidines, shown

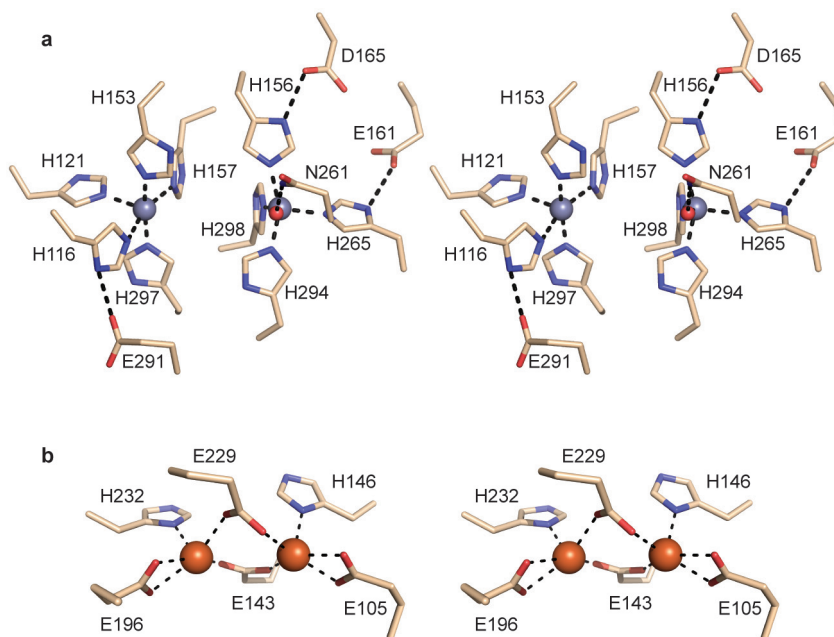
with the weighted $2F_o - F_c$ density contoured at 2σ . **c–e**, The dimetal centre superposed with the anomalous difference map, contoured at 5σ (**c**), the $F_o - F_c$ density calculated with the zinc atoms omitted, contoured at 3σ (**d**), and the $F_o - F_c$ density calculated with the ordered water molecule between M1 and Asn261 omitted, contoured at 3σ (**e**).



Extended Data Figure 6 | Western blot analysis of SCD expression.

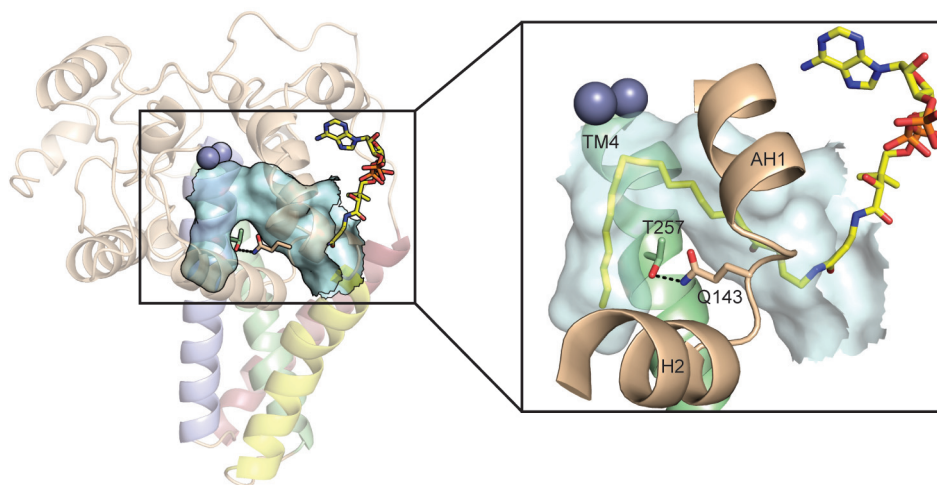
a, b, Analysis of two separate yeast expression trials after introduction of mutations to mouse SCD3 to impart catalytic specificity of mouse SCD1. Contents of lanes are as indicated in the gel. The position of SCD is indicated by a black star. Additional bands are other proteins detected by the polyclonal

antibody. Dotted line in **a** shows the portion of the complete gel image included in Fig. 2f; dotted line in **b** shows the corresponding expression trials from the second experiment. **a,** Expression trial 1, with gel artefacts in lanes 2 and 3. **b,** Expression trial 2, with gel artefacts in lanes 4, 6 and 7.



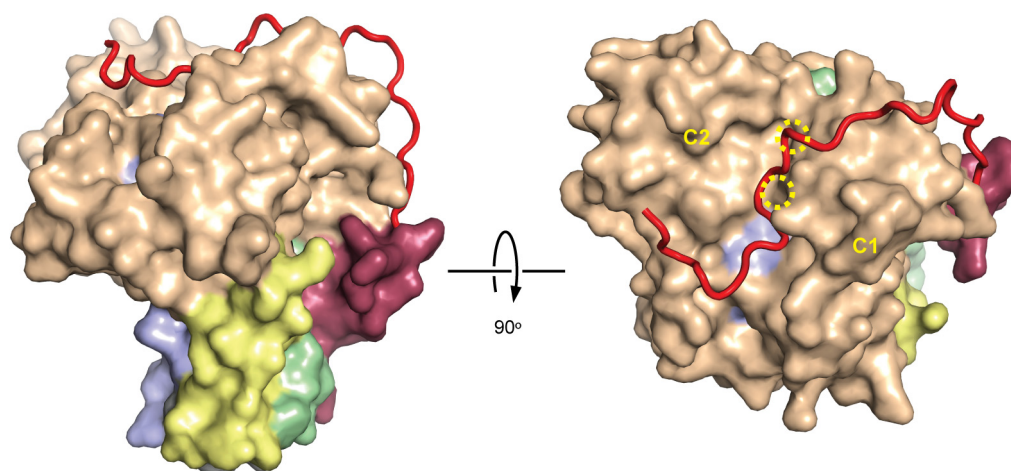
Extended Data Figure 7 | Coordination in diiron-containing desaturases.
a, Stereoview of residues forming both the first and second coordination shell around the dimetal centre in mouse SCD1. **b**, Stereoview of the coordination of

the dimetal centre in the stearoyl-acyl carrier protein desaturase from the castor bean (PDB accession 1AFR).



Extended Data Figure 8 | The Thr257–Gln143 hydrogen bond blocks product egress. The surface of the substrate tunnel housing the acyl chain is shown, with the structural elements AH1, H2 and TM4, and the hydrogen-bonded residues Thr257 and Gln143 highlighted in the inset. The proximity of

these two residues creates the kinked shape of the substrate tunnel, and their separation would result in a larger opening capable of releasing the product into the bilayer.



Extended Data Figure 9 | The SCD1 N terminus. Two perpendicular views of mouse SCD1, from within the plane of the membrane and from the cytoplasmic side, showing the interaction between the N terminus (red ribbon) and the

cytosolic domain (beige surface). The dashed yellow circles indicate the approximate location of the metal atoms.

Extended Data Table 1 | Crystallographic and structure refinement statistics

	Phasing Data Set	Refinement Data Set
Data collection		
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit Cell (Å)	a=77.51, b=114.53, c=140.97	a=77.06, b=113.77, c=141.70
Wavelength (Å)	1.2541	0.9795
Resolution (Å)	38.8 - 2.8	47.4 - 2.6
R _{sym}	0.18 (1.17)	0.11 (0.53)
I/σI	9.5 (1.3)	14.4 (3.1)
Completeness (%)	99.1 (94.0)	98.0 (94.0)
Redundancy	7.9 (7.5)	6.2 (5.0)
Refinement		
Resolution (Å)		47.4 – 2.6
No. reflections		38016
R _{work} /R _{free}		20.3 / 23.5
No. atoms		
Protein		5231
Ligand/ion		188
Water		86
B-factors (Å ²)		
Protein		46.8
Ligand/ion		51.5
Water		40.1
R.m.s deviations		
Bond lengths (Å)		0.002
Bond angles (°)		0.633
Ramachandran Plot		
Favored (%)		93.8
Allowed (%)		6.2
Outliers (%)		0.0

CAREERS

MENTORS Nobel prizewinners recall advice from their early days go.nature.com/ffqgq7

SCIENCE COMMUNICATION Keep it simple and read your audience go.nature.com/nkuzus

NATUREJOBS For the latest career listings and advice www.naturejobs.com

DONAT SOROKIN/TASS/CORBIS



The rise of blockbuster biological drugs should translate into hiring opportunities for early-career scientists.

BIOMEDICAL RESEARCH

Drug hunters wanted

The biotherapeutics industry is burgeoning — and it needs scientists with specialized disease knowledge and technical savvy to join in the drug-discovery efforts.

BY JEFFREY M. PERKEL

A newly approved class of anti-cholesterol medications could be the latest in a long line of 'biopharmaceutical blockbusters'. These drugs not only produce big revenue for pharmaceutical companies, but also represent employment opportunities for early-career scientists who want to develop cutting-edge therapies. To get into the game, aspiring young researchers must tailor their training and skills to the industry.

Biopharmaceuticals — or 'biologics' — are complex drugs that are manufactured or extracted generally from biological sources. They include proteins produced in engineered cells, other large molecules and live cells. A 2013 report by the Pharmaceutical Research and Manufacturers of America, a trade association in Washington DC, said that 907 biologics were in development in the United States alone (*Medicines in Development: Biologics* PhRMA; 2013). And drug manufacturers worldwide are keen to expand the field.

In July, the US Food and Drug Administration (FDA) approved a cholesterol-lowering biologic called alirocumab. A similar compound, evolocumab, could get full approval later this year; it passed the FDA's preliminary regulatory-assessment panel in June. The drugs are the first candidates in their class — known as PCSK9 inhibitors, after the protein whose function they block — to advance this far in any drug company's pipeline. Alirocumab and evolocumab both belong to a group of medications called monoclonal-antibody therapeutics, which ►

► work by binding to and altering target molecules; PCSK9 inhibitors specifically seek out and inactivate a protein in the liver, leading to a decrease in the amount of low-density lipoprotein (LDL) cholesterol in the bloodstream.

Some analysts project multibillion-dollar sales for these two cholesterol combatants. The 20 highest-grossing biologics in 2013 generated more than US\$1 billion each, and the top three collectively raked in more than \$28 billion (*Nature Biotechnol.* **32**, 992–1000; 2014). “I think it’s fair to say we expect to see strong growth in biotherapeutics,” says Raymond Amato, who oversees hiring for the research and development division of pharmaceutical company Pfizer in Groton, Connecticut.

BREAK INTO BIOLOGICS

Early-career researchers who hope to get a toe in the door need detailed knowledge of relevant areas, such as neurobiology, cardiovascular disease or immunology, and should be deeply familiar with specialized techniques such as antibody engineering, next-generation DNA sequencing, bioinformatics or genetic manipulation. It also helps to have proven problem-solving abilities, creativity and leadership skills — and the capacity to work in a team, says Mark Kowala, chief scientific officer of cardiometabolic and diabetic complications at Eli Lilly, a health-care company in Indianapolis, Indiana.

Lilly alone has about 100 job openings in therapeutic development, including biologics. In the biopharma arena, the company’s research and development positions are wide-ranging and include lab technicians, senior scientists, statisticians, pharmacologists and toxicologists, says Jennifer Porath, a human-resources director for global recruiting at Lilly. A PhD is not necessarily required, she adds. Kowala’s department, for example, has two lab heads without PhDs, who represent about one-tenth of the group. Competition for these jobs is fierce. Paul Nioi, a translational biologist at Amgen in Cambridge, Massachusetts, says that his most recent postings attracted about 50 applicants each over the first week. At Pfizer, advertisements for entry- to mid-level-research positions draw 150 to 250 candidates apiece, Amato says.

So how can early-career researchers boost their viability as candidates? In addition to fundamental-biology knowledge and lab skills, it helps to gain familiarity with concepts important to the biopharmaceutical business, such as the clinical development and large-scale production of monoclonal antibodies, recombinant proteins and cellular therapies. Those who hope to work for specific companies should

“They were a little surprised that I was thinking about the disease relevance of my work.”



When it comes to nabbing a job in the biopharmaceutical industry, ‘soft’ skills matter — and among the most important is the ability to communicate. As in all scientific disciplines, scientists in biopharmaceutical research and development must write papers and give presentations. But they must also communicate ideas to company executives and colleagues and pitch a message effectively to both audiences. “It’s sometimes like being a used-car salesman, as well as being a scientist,” says Andrew Adams, a senior research scientist at the pharmaceutical company Eli Lilly in Indianapolis, Indiana. “You have to be really persuasive to people who sometimes will be at odds with the idea that you’re putting forward.”

Diana Ritz, a principal scientist who works for GlaxoSmithKline (GSK) in Upper

seek out knowledge relevant to the firms’ drug-development targets. Amato says that Pfizer, for example, in collaboration with the gene-editing firm Editas in Paris, has begun to pursue cell-based immunotherapies. Among the technical skills required for those jobs are primary T-cell culture and experience with cell manufacturing, characterization and quality control.

John Beals, a protein chemist who advises on bioproduction at Lilly, hires engineers who can clone and express genes, measure protein properties and assess molecular structure. Pharmaceutical giant GlaxoSmithKline (GSK) looks for scientists who are skilled in analytical chemistry, cell biology and protein purification, says Joseph Tarnowski, a senior vice-president at GSK’s office in Upper Merion, Pennsylvania.

A skill set that was directly translatable to biologics development proved invaluable for Wei Ni, a research scientist who joined Lilly straight from her postdoc at the University of California, San Francisco. Her expertise in pharmacology and molecular biology helped to set her apart from her competitors, says Kowala, who hired her for his drug-hunting team.

Ni also demonstrated several crucial ‘soft’

SOFT TOUCH

More than lab savvy

Merion, Pennsylvania, can attest to the importance of communication skills. She was hired directly from a PhD programme in chemical engineering, and notes that the doctorate taught her how to make a research project particularly attractive to biopharma companies and how to frame it for industry eyes. Presenting talks over the years also helped her to hone her message, she says. “It’s a matter of focusing on things like increases in productivity, time and cost savings versus mechanistic minutiae that might get academics going, but don’t impact the bottom line at all,” she says.

Being able to think outside the box is a plus, says Adams. “Can you make mediocre ideas into good ideas by applying creative thinking?” he asks. Joseph Tarnowski, a senior vice-president at GSK, recommends that candidates’ applications include an example of an unconventional project or strategy, to show that an applicant is curious and innovative. “That would get our attention,” he says.

Other key attributes include self-motivation, enthusiasm, the ability to think on one’s feet — and a genial personality. “You’re going to spend a lot of time with them; you have very challenging assignments,” says John Beals, a team leader at Lilly. “If you are a joy to work with and people like you, then you get a ripple effect of better relationships and better problem-solving.” **J.M.P.**

skills (see ‘More than lab savvy’). Lilly asks candidates to give a research seminar during their job interviews. Ni took the opportunity to show that she could think both broadly and creatively, and communicate her ideas well. “My first two slides were on epidemiology and clinical trials that were relevant to my research,” she says. “I think they were a little surprised that I was thinking about the disease relevance of my work.” She also demonstrated an ability to develop alternate ideas when her initial hypothesis was proved wrong, and to find ways to fill technical gaps — in this case, quantifying biological phenomena from images — in her expertise.

What matters most, says Stan Crooke, chief executive officer at Isis Pharmaceuticals in Carlsbad, California, is what a scientist does when handed full responsibility for a research project and concomitant resources. “Decide what area of science you’re interested in, demonstrate the highest-quality performance, and that,” he says, “is the answer.” ■

Jeffrey M. Perkel is a freelance writer in Pocatello, Idaho.

UNINHABITABLE ZONE

Approach with extreme caution.

BY IAN STEWART

The primary gleamed benignly against a background dusted with stars. From the inner system it would have blazed angrily in the sky, but this planet was sufficiently distant to be safe.

"Perfect!" Unit-Peripheral declaimed. "I find it hard to believe that such a beautiful, balmy world can be lifeless — yet it is! I will authorize colonization immediately."

"The world is *currently* lifeless," Plug-in-43 pointed out. "The sondage team suspects that living creatures were once present here. They left ... traces. Indicative of bipedality."

"Garbage! Nothing can locomote on two appendages! The arrangement is unstable. The marks were made by a meteoroid impact."

"Perhaps. The sondage team has formulated an alternative hypothesis. It is ... strange."

"It would have to be. What is it?"

"Extremophiles."

Unit-Peripheral paused to recompute and got the same result. "Nonsense. Extremophiles occupy *extreme* habitats."

"Extremism is relative, Unit-Peripheral. These beings made a brief visit, decided this world was not to their liking, and departed."

"Where to?"

"The sondage team suspects ... absurd, I know, but they take their duty to be open-minded terribly seriously ... that the extremophiles departed for the inner system."

Unit-Peripheral expressed indicatory signals in the direction of the severely overheated local star. "Into *that*?"

Plug-in-43 said nothing, but q-mailed an image: two parallel linear series of depressions, melted into the rock.

"The inference of warmlife is highly improbable," Unit-Peripheral stated. "This system's gas and ice giants are far too close to the star to be habitable. The cryonic zone starts with the ice dwarves, of which this is the innermost. Where does the sondage team imagine these extremophiles reside? On a comet?"

Plug-in-43 q-mailed another image. "Barrenworld-3. It resembles Fumarole in our own system."

"Where rocks run molten, forming magma oceans."

"As they do on Barrenworld-3. The sondage team suspects this system has evolved extremophiles that can inhabit such worlds. In fact, they suggest these organisms might largely *consist* of molten rock." Noticing Unit-Peripheral's thunderous mode-expression, Plug-in-43 hastened to expand. "Held in ... a

high multitude of ... tiny cocoons ... able to stand the searing tempera —"

"*Cluster*glitch! Even extremophiles can't be *that* extreme! The radiation alone —" "Barrenworld-3 has a protective magnetic field."

Unit-Peripheral came to a reluctant decision. "Protocol requires me to act as if I take these insane conjectures seriously. Dispatch six-by-six hardened probes."

The crew descended to the planet's orange and brown surface for a refreshing sunbath, and waited. Eventually one heavily damaged probe struggled back. Its report was disturbing. Not only a gaseous atmosphere, but —

"Oxygen?"

"Six-plus-one six-by-sixths by volume."

"And the cyanic areas truly *are* liquid rock?"

"Molten dihydrogen monoxide, Unit-Peripheral."

"Despite vast amounts of atmospheric dioxygen carbide, indicative of past carbida-tion events on a huge scale, there remains *free* oxygen? Why did such a corrosive substance not combine with other elements long ago?"

"It did. It still does. But the poison is regenerated."

Unit-Peripheral emitted a warning flash. "This smacks of theory-saving! *How*?"

"By extremophiles."

The flash shorted out, destroying a food-booth. "Whenever something impossible is proposed, it is justified by invoking extremophiles! I've had enough of this nonsense. I cannot accept that intelligent space-traversing creatures can regenerate poison gas."

"Apologies. I was unclear. The poison is regenerated by a different extremophile."

Unit-Peripheral vibrated in a mix of anger and terror. "There's *more than one* of them?"

"Millions of formats, all multiple upgrades."

"Even with a magnetic field, these creatures must live underground."

"No, they lie around the edge of the magma and ... *sunbathe*. Like us."

"At least they have the sense to stay out of the liquid rock."

"No, they periodically immerse themselves in it. To — uh — I know no other way to say this, Unit-Peripheral. To keep cool."

"Cool? The probe is suffering from an

overactive imagination. Oxygen is one of the most corrosive gases known to cryokind! How do these creatures avoid reacting with it?"

"They don't. They use it as an energy source."

"But surely the organisms *themselves* would oxidize!"

Plug-in-43 abased its bias currents as a submission gesture.

This one would not be well received. "Sometimes they do. The probe saw localized conflagrations in which thousands of extremophiles died."

"The bipedal ones?"

"No. Sessile colonies."

The team names them 'fire-forests'. They cannot reproduce unless there is a conflagration."

"They have babies by setting themselves on fire?"

"Only some of the sessile formats. Other formats use different methods — all bizarre."

"They don't flake off circuit copies like we do?"

"No, they ... never mind, you wouldn't believe it if I told you."

"Do the bipeds set themselves on fire?"

"No, not deliberately. However, they do derive their energy from millions of diminutive internal fires —"

"*Enough!* Another word and you will be demoted to long-term storage! I am stacked up to the back slots with this arrant nonsense! Put out a general order."

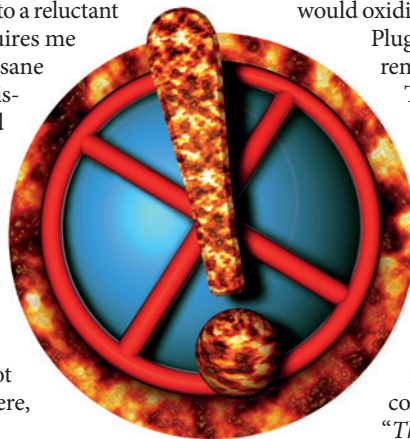
"We are depositing a colony as planned? The conditions here are ideal for our super-conductive brains, and —"

"No! We are departing forthwith. We will report that this system has no habitable worlds. Which is true. For if any of us were to live here for more than a few cycles, the mere thought of those *things* cavorting in magma would drive us mad!"

Unit-Peripheral's sensorium swivelled back to the q-mailed image: a multicoloured globe with ugly patterns of toxic cyan, sickly green patches, deathly brown scars, all overlain with ghastly albescent tendrils. It paused to purge its processors, and uttered its final words on the topic, dripping with scorn. "Why couldn't they have been *normal*, like us?" ■

Ian Stewart, emeritus professor at the University of Warwick, writes popular science books and science fiction.

ILLUSTRATION BY JACEY



naturejobs

CAREER

GUIDE

ASIA-PACIFIC

NATURE, VOL. 524, NO. 7564 (13 AUGUST, 2015)

COVER IMAGE: DANNY HU/GETTY

As long ago as 1803, Napoleon Bonaparte apparently pointed to China on a map and warned: “Here lies a sleeping giant. Let him sleep, for when he wakes up, he will shock the world.”

These days, as the world has seen, the giant is well and truly awake, not least when it comes to science. China has almost doubled the proportion of its GDP spent on R&D over the last decade, and in 2013 overtook the combined R&D spending of the 28 nations of the European Union. The Organisation for Economic Co-operation and Development predicts it will leapfrog the US by 2020.

For anyone considering their career in science, China's rise is reason enough to think about a move to Asia, but it is not the only one. A little to the east, South Korea is second only to Israel in the proportion of its GDP it spends on R&D, and Japan is not far behind. Singapore has built up its research and innovation capacities rapidly since the turn of the century by luring foreign talent with offers of large salaries. Australia and New Zealand are playing to their strengths by focusing limited resources on the fields in which they excel.

Researchers in the other parts of the world that have traditionally dominated the scientific playing field are facing tough times as competition for jobs and grants increases, and R&D budgets shrink or flatline. The UK, for example, spent 1.6% of GDP on R&D in 2013 – less than it did a decade earlier. US share of global R&D spending fell from 34% to 30% in the decade to 2011.

Hardly surprising then that many ambitious scientists are embracing opportunities in the Asia-Pacific region. Of course, taking a job far from home is a big leap, so in this Naturejobs Career Guide we present first-hand accounts of what it's like to work in this region along with tips from employers and key facts and figures.

Most people who take up overseas opportunities have positive, career and life-changing experiences. For a minority it doesn't work out because of unexpected challenges. Those of us who have put this guide together hope it increases the likelihood that readers considering taking the plunge will be glad they did.

Nic Fleming
Guest Editor

CONTENTS

S2 OVERVIEW

A graphic picture of opportunities and achievement across Asia Pacific

S6 CHINA

Spending big on R&D to claim a place at the top table of scientific research.

S16 JAPAN

Teamwork and hierarchy hold the key to success in state-of-the-art laboratories.

S20 AUSTRALIA

High-quality climate-change research a priority, but funding constraints bite.

S24 SINGAPORE

Science funding plentiful for an ambitious and international workforce.

S26 SOUTH KOREA

Investment and recruitment drive fuel programme to build a creative economy.

S28 NEW ZEALAND

A natural magnet for geoscientists, but post-doc opportunities are limited.

EDITORIAL: Nic Fleming, Stephen Pincock, Larissa Kogleck, Rebecca Dargie, Nisha Gaind, Anna Novitsky, Julie Gould, Smriti Mallapaty, Karen McGhee, Bianca Nogrady, Sarah O'Meara, Mark Zastrow. **ART & DESIGN:** Alisdair Macdonald, Wesley Fernandes, Kate Duncan, Chris Gilloch. **PRODUCTION:** Sue Gray, Karl Smart, Ian Pope, Robert Sullivan, Chandler Gibbons. **MARKETING:** Claire Jones. **SALES:** Janet Cen, Yuki Fujiwara, Helen Hill, Yuko Takai, Nicole Yu. **GLOBAL HEAD OF CAREER SERVICES:** Nils Moeller. **PROJECT MANAGER:** Anastasia Panoutsou. **ART DIRECTOR:** Kelly Buckheit Krause. **PUBLISHER:** Richard Hughes.

NATUREJOBS CAREER GUIDE

The Naturejobs Career Guide Asia Pacific 2015, a supplement to Nature, is produced by Nature Publishing Group, a division of Macmillan Publishers Ltd. Naturejobs (naturejobs.com) is the global jobs board and career resource for scientists.

Nature Editorial Offices
The Macmillan Building
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0)20 7833 4000
Fax: +44 (0)20 7843 4596/7

CUSTOMER SERVICES

To advertise with Naturejobs, please visit naturejobs.com/support or email naturejobs@nature.com
feedback@nature.com
Copyright © 2015 Nature Publishing Group.
All rights reserved.

RIISING IN THE EAST

When weighing up a country as a potential place to live and work, it is worth considering how it compares with the rest of the region for science output, government funding and general cost and quality of life.

HOW SCIENCE IS SPREAD

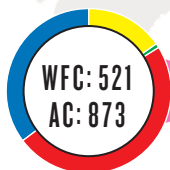
There are both giants and minnows in the Asia-Pacific region when it comes to the number of academic papers published in Nature Index, a database that tracks the national affiliations of articles published in high-quality journals. This is clear in terms of both raw article count (AC) and adjusted weighted fractional count (WFC). China's emphasis on chemistry and New Zealand's focus on earth and environmental sciences are clear. See below left for a detailed description of the Nature Index.

Bubbles are sized according to WFC. The pie charts represent the countries' output in the categories below.

- Life sciences
- Earth & environmental sciences
- Chemistry
- Physics

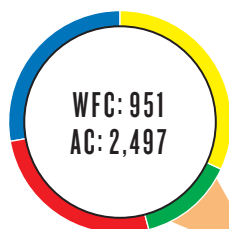
*Each slice represents the proportion that each subject area contributes to an institution's overall WFC. Subject areas can overlap, so the total percentage may exceed 100%.

At just under 5.5 million, Singapore's population is the smallest among the six countries. If WFC were calculated per capita, Singapore would come out on top by far.



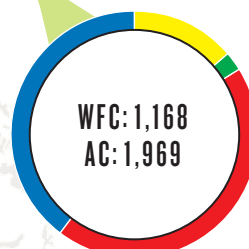
SINGAPORE

Australia leads the way in life sciences in the region, with 35% of its WFC attributed to the field.



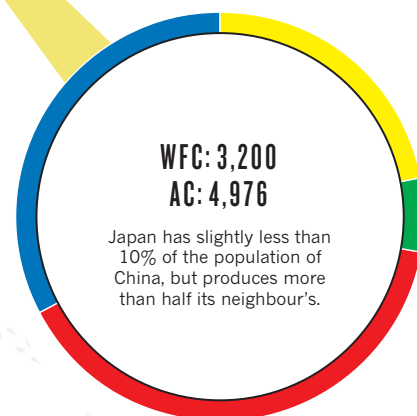
AUSTRALIA

The physical sciences contribute 46% of South Korea's WFC — the highest proportion in the six countries.



SOUTH KOREA

JAPAN



Japan has slightly less than 10% of the population of China, but produces more than half its neighbour's.

WFC: 96
AC: 275

NEW ZEALAND

New Zealand has by far the lowest article output in the region, but is unique in having more than one-third — 37% — of its WFC in Earth and environmental sciences.

NATURE INDEX

The Nature Index database tracks the affiliations of high-quality scientific articles, and charts publication productivity for institutions and countries. Article count (AC) includes the total number of affiliated articles. Weighted fractional count (WFC) accounts for the relative contribution of each author to an article and applies a weighting to correct imbalances in the Index's subject coverage. This Career Guide draws on Nature Index data derived from articles published in calendar year 2014. WFC is used throughout as the primary metric, because it provides an even basis for comparison. For more information, visit natureindex.com/faq

KEY

CHINA

JAPAN

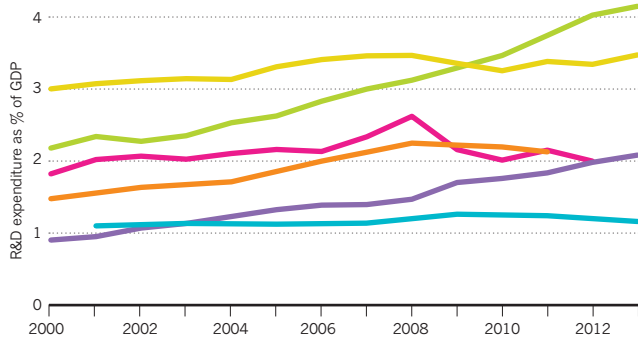
AUSTRALIA

SINGAPORE

SOUTH KOREA

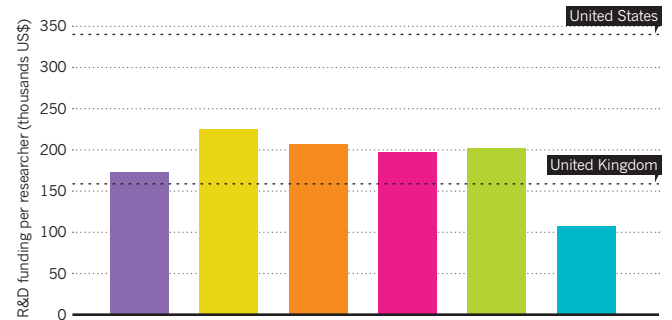
NEW ZEALAND

FUNDING OVER TIME



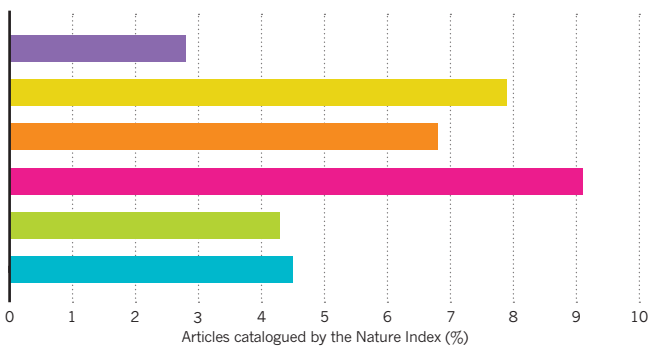
Research and development (R&D) spending as a proportion of gross domestic product (GDP) has risen in all six countries since 2000. The increases in have been largest in China and South Korea, both of which have almost doubled their investment in R&D in the past decade.

SPENDING PER RESEARCHER



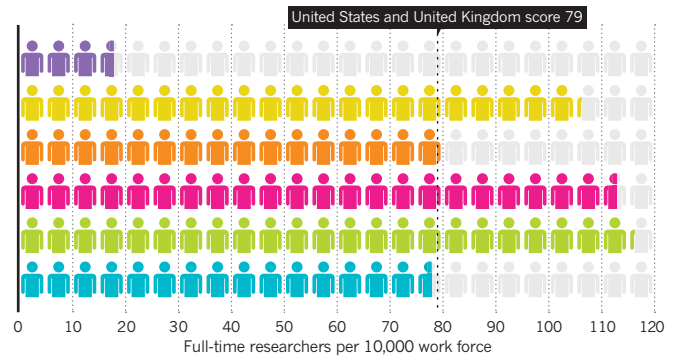
Japan spends the most per researcher, and New Zealand spends by far the least. Figures are the most recent available for each country, and are normalized for purchasing power.

RESEARCH QUALITY



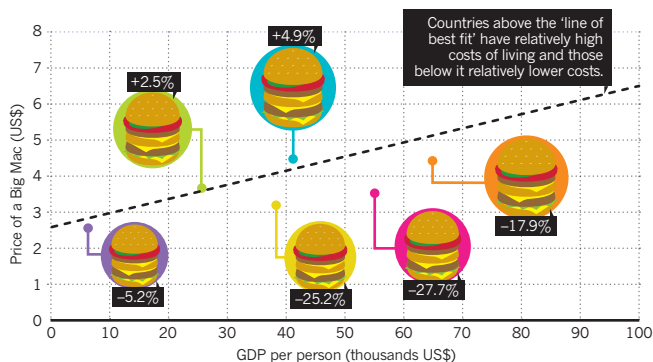
The ratio of each country's article output in the Nature Index to its biological and physical sciences output in the Scopus database is an indicator of how much of its science is published in the journals most favoured by researchers.

RESEARCHER DENSITY



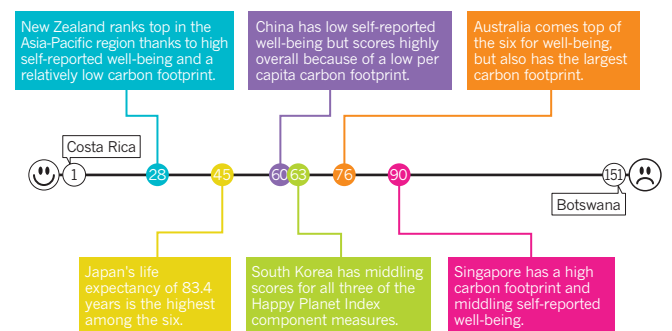
South Korea is the region's leader in the proportion of its workforce doing research, and Singapore is not far behind. With 1.4 million researchers in a population of 1.37 billion, it will be some time before China catches up. Figures are the most recently available for each country.

BIG MAC INDEX



The Big Mac Index was created by *The Economist* as a light-hearted illustration of the idea that over the long-term, exchange rates should move so that the prices of identical goods equalize in any two countries. Taking into account varying labour costs, the dotted black line on the graph is the 'line of best fit' for Big Mac prices plotted against GDP per person for 48 countries (plus the euro area). Assuming that Big Mac prices are at least loosely representative of the prices of other goods, countries with average costs of living would lie on the line.

HAPPY PLANET INDEX



Produced by UK think tank the New Economics Foundation, the Happy Planet Index (HPI) assesses how well countries provide inhabitants with long, happy and sustainable lives. It ranks countries using a formula that takes into account self-reported well-being, life expectancy and the WWF Ecological Footprint measure of the per capita amount of land of average productive biocapacity required to sustain national resource consumption. Numbers shown on the axis are the countries' 2012 HPI rankings out of the 151 nations included.

CHINA

With a declared aim to become a global science leader by 2050, the country's unprecedented research spending splurge is creating a wealth of study and work opportunities.

BRUCE YUANYUE BI/GETTY



Shanghai, with its mix of historic and modern architecture, hosts the regional headquarters of many multinational companies.

Within six months of moving to Shanghai, physics postdoc Kevin Huang was heading back to the United States. Far from returning with his tail between his legs, Huang had found several funding sources for a new project during his first month at Fudan University. He was making the trip to put together the ambitious collaboration with the Los Alamos National Laboratory, New Mexico, and the National High Magnetic Field Laboratory in Tallahassee, Florida.

“Fudan University is building new chemistry, physics, and environmental engineering buildings in stark contrast to the institutions in the US where funding seems to be slashed every year or stagnant at best,” says Huang, whose parents immigrated from China to the US in 1983.

China is set to become the world's largest R&D investor by the end of the decade, opening job opportunities and driving up standards, says Lei Jiang from the Institute of Chemistry, Chinese Academy of Sciences. He sits on a selection committee for the 1000 Youth Talent programme — a scheme that offers good salaries, funding and welfare benefits to high-flying Chinese scientists aged under 40 who take positions in the country.

Jiang says competition for places has become intense since the scheme launched in 2011 — part of the broader 1000 Talents drive — which began three years earlier to attract some 2,000 leading scientists, entrepreneurs and finance experts of any nationality within 10 years.

“This year about 700 candidates were

selected from more than 2,300 applications. The research quality was high, so it was difficult to eliminate anyone. More than 70% of them could get academic positions in the US. This is completely different from 15 years ago.”

China's stated aim is to become a global leader in innovation by 2050. It has almost 3,000 higher education institutions, and the number of university applicants has jumped from 5.3 million in 2002 to 9.5 million in 2015.

‘IMAGINE SILICON VALLEY IN THE EARLY DAYS — THIS IS WHAT IS HAPPENING AROUND ME.’

The government has committed large sums to high-profile projects such as thorium-based nuclear power plants, as well as basic research spending, which has historically received less funding than in other developed countries.

Researchers seeking funding can apply to national agencies such as the National Natural Science Foundation of China, Ministry of Science and Technology, and other bodies, once a year. Smaller sums can be applied for directly through their websites, while larger grants require a presentation in front of committee members. There are also programmes run by local governments.

For scientists looking to do research in the commercial sector, China is enjoying strong

private sector growth. Firms range from well-known multinational pharmaceutical brands such as GlaxoSmithKline, and home-grown equivalents like Pharmaron, to innovative Chinese start-ups such as fast-growing biomedical firm WuXi AppTec. Management consultants McKinsey predict that by 2016 more than four out of five global life sciences organisations will be conducting R&D in China.

Chuyao Peng received a PhD offer from Oxford University to research condensed matter physics, but chose to join Chinese tech start-up Shanghai Superconductor Technologies. “I’m really excited about the opportunities for young scientists here. Imagine Silicon Valley in the early days — this is what is happening around me.”

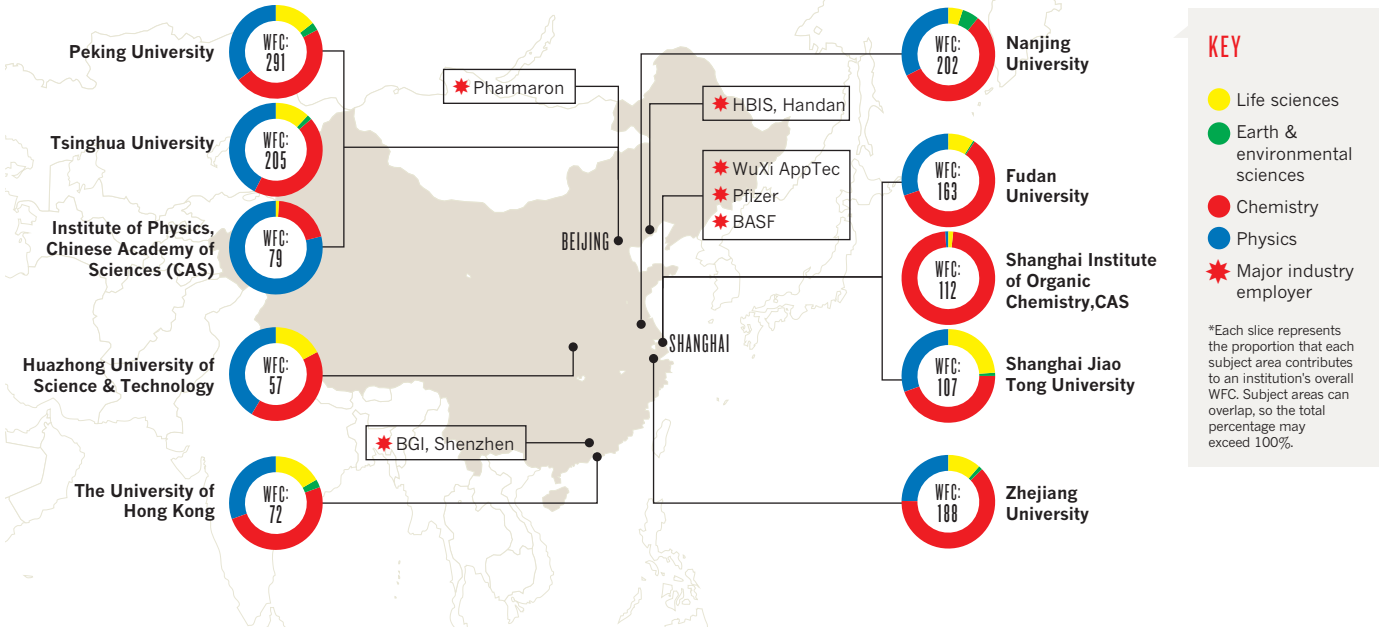
But despite China's fast growth and modernisation, the country's tightly controlled systems can still be challenging for foreigners or returnees.

Faxian Xiu of Fudan University's Department of Physics points out that the bureaucracy at institutions is irritating: “If you need new equipment, for example, the extensive paperwork needs to be done well ahead of time. It can take up to half a year to arrive.”

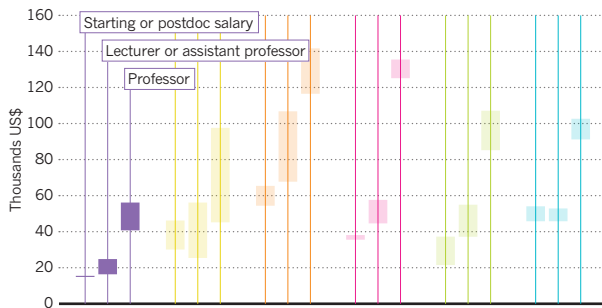
Internet restrictions in the workplace are also a major headache. Even opening emails can take a long time and the blocking of many websites is frustrating. Other challenges include chronic air pollution, poor water quality and high population densities in urban centres. ■

WHERE TO WORK

The below charts represent the research output included in the 2014 Nature Index for ten of China's leading institutions, and the contributions of different subjects, measured by weighted fractional count (WFC).

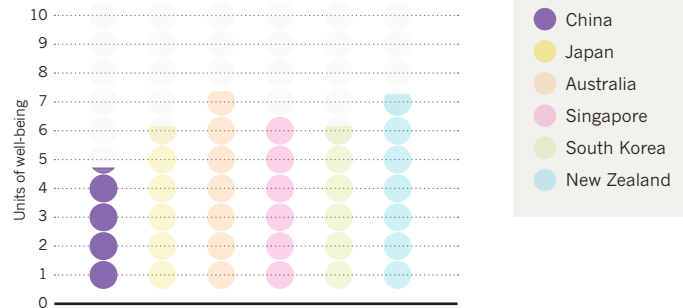


SALARIES



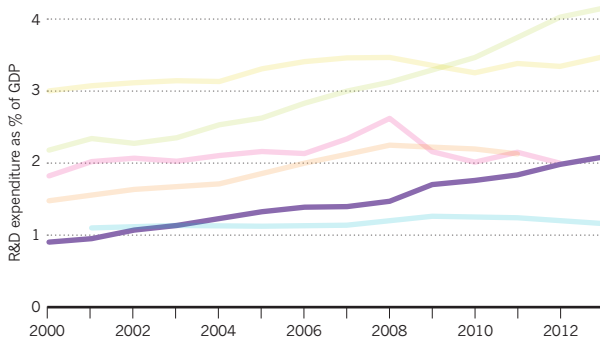
Salaries in China are the lowest of the six profiled Asia-Pacific countries, however the cost of living is also relatively low. Ranges are on the basis of reported data.

REPORTED WELL-BEING



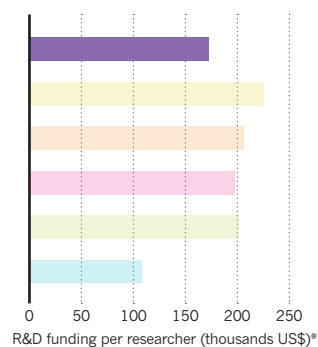
The Chinese report the lowest levels of life satisfaction among residents of the six countries.

FUNDING OVER TIME



China has been the most consistent in its annual increases in the proportion of GDP spent on R&D since over the last decade.

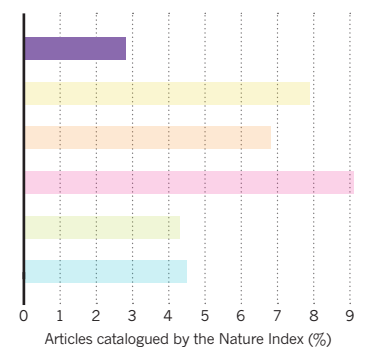
SPENDING PER RESEARCHER



Only New Zealand spends less per researcher of the six nations, however China is catching up rapidly.

*Figures are normalized for purchasing power, and are the latest available for each country.

RESEARCH QUALITY



The proportion of China's research output published in journals considered by the Nature Index is the lowest of the six.

**BICHENG YANG****>EMPLOYER****>DIRECTOR OF COMMUNICATION
AND PUBLIC ENGAGEMENT****>BEIJING GENOMICS INSTITUTE (BGI)****What do you look for in candidates
for research positions?**

We value innovation most. Science is never set in stone but needs new thoughts and ideas. Moreover, in an era of big data there are lots of breakthroughs to be made. We also highly value those who have an open mind when it comes to knowledge and technologies.

**Are your pay rates equivalent to
those in the West?**

Salaries in China are becoming more competitive. Companies have begun offering enhanced packages for promising positions with good career prospects to attract academic talents from across the world.

**How could a move to China benefit a
researcher's career right now?**

China's profile and stature is rising in many fields of research. Scientists benefit from abundant resources and incentives provided by national and local government to encourage research and development. For example, in 2011 in Shenzhen, where BGI's headquarters is located, the Shenzhen Municipal People's Government set out its twelfth 'five-year plan' to support research and innovation within six strategic emerging industries - biotechnology, internet, renewable energy, advanced materials, cultural creativity and information technology. Therefore, with China's encouraging policies, researchers will have more resources and funding, a broad platform, and more professional opportunities.

**Is BGI planning high levels of recruitment
in the immediate future?**

The future of bioscience - and BGI - will require multidisciplinary research talents from the fields of physics, computer science, mathematics, biology and beyond. As talent is always the source of innovation, we have already recruited many scientists trained abroad, and intend to recruit more. ■

CHINA RISING

The rise and rise of Chinese R&D spending 2007-08 global economic slowdown saw banks go to the wall, property bubbles pop, consumers tighten their belts and tax revenues diminish. It was not long before research in both the public and private sectors was also cut back as science budgets shrank in most countries — with one notable exception.

As shockwaves were felt across the globe in 2009, China's research sector saw a 28% increase in R&D funding on the previous year to reach 1.7% of GDP. With domestic growth still strong, Chinese leaders were undeterred from using the country's financial muscle to quickly turn it from a scientific backwater to a world leader.

An increase in annual R&D growth of an average of 23% a year over the past decade has already seen China overtake the European Union in the spending race. In its *Science Technology and Industry Outlook 2014* report, the Organisation for Economic Co-operation and Development predicts it will take the top investment position from the US by 2020.

The results of the splurge are evident. China has become the second most prolific producer of peer-reviewed research articles according to the Royal Society, and has taken a global lead in the number of patent applications filed each year, data from Thomson Reuters show.

It has also launched a series of major R&D-based projects such as a space station and the China brain project, dedicated to research into artificial intelligence and neurological diseases.

"In some fields [China] is at the frontier of technological knowledge, and the growth of published research is extraordinary," writes Geoff Mulgan, chief executive of the UK's National Endowment for Science, Technology and the Arts (Nesta) in the organisation's 2013 report *China's Absorptive State*.

Historically, scientists' salaries and bonuses have been linked to the number of articles they publish in high-impact journals. Materials scientist Anthony Cheetham, vice-president of the Royal Society, noted in a recent interview for *Nature* that this professional emphasis can

occasionally 'drive some bad behaviour' such as the falsification of data and plagiarism.

Steffen Duhm, a German physics professor who joined Soochow University in Suzhou, Jiangsu, three years ago, agrees that the publication culture can stifle exploration. "My students expect from me that I know (and will tell them) the results of an experiment before they perform it."

Likewise, the government set high targets for invention patents in its 15-year plan to achieve 3.3 patents per 10,000 people by 2015 — a figure it looks set to surpass.

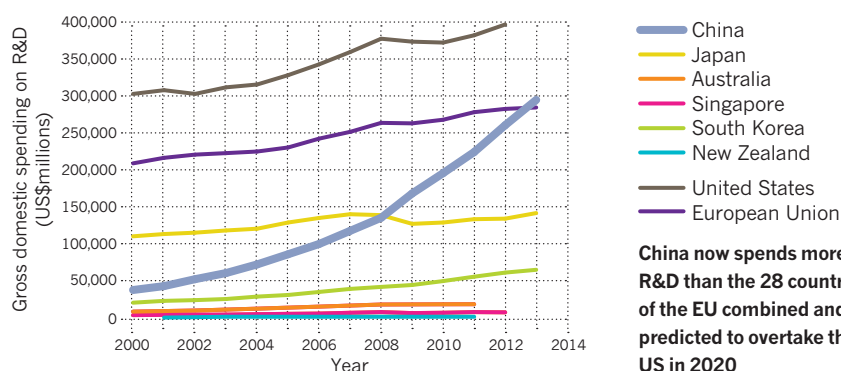
However the quality of some of these patents has been called into question. *The Economist* noted that just 5% of the patents filed in China in 2013 were registered globally, rather than just locally — an indication of their perceived low value — suggesting that inventors are mainly striving to meet bureaucratic targets. Others say that despite this, producing commercially viable innovations is a priority for authorities.

CHINA PUBLISHED A 15-YEAR PLAN TO TRANSFORM THE COUNTRY INTO A SCIENCE AND TECHNOLOGY POWERHOUSE BY 2050

The country's scientific achievements are inevitably part of its long-term strategic goals. It aims to raise its R&D spend to 2.5% of GDP by 2020 — up from 1% in 2000. Meanwhile the US share of global R&D spending has fallen from 34% to 30% 2001-2011. In 2006, the Chinese government published a 15-year plan which included plans to transform the nation into a global science and technology leader by 2050. Assuming it is able to maintain the extraordinary rates of growth in R&D spending seen in recent years, few would bet against it achieving this goal.

None of its economic rivals can match these levels of growth. ■

THE RISE AND RISE OF CHINESE R&D SPENDING



China now spends more on R&D than the 28 countries of the EU combined and is predicted to overtake the US in 2020

SOURCE: OECD

ENTRY REQUIREMENTS

The employers of foreign nationals must apply for work permits on their behalf. As in most countries a case must be made that there are no nationals capable of fulfilling the requirements of the job. Potential employees' CVs, university transcripts and certificates are presented to local authorities.

Once permission to apply for a work permit has been granted, individuals must apply for an employment Z visa for a single entry into China. After arriving in the country, the employer and their employee have one month to apply for an official residence visa and work permit. This can take up to three months to process and is valid for one year after which it must be renewed annually.

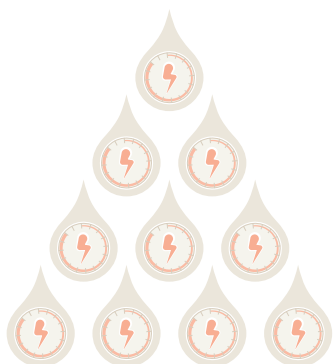
ACADEMIC YEAR

Autumn term runs from September to late January or early February, ending by the **Spring Festival** or **Chinese New Year**.

Spring term runs from late February or early March through to late June or early July. **Summer break** is from July to the end of August.

OPPORTUNITIES & CONTACTS

1,000 Talent Plan of Foreign Experts, also known as the National Recruitment Program of Global Experts, and the 1,000 Youth Talent Plan: <http://onestop.globaltimes.cn/what-is-the-1000-talent-plan-for-people-with-chinese-heritage-and-how-does-it-work/>
100 Talents Program for overseas Chinese applicants: search@ucas.ac.cn
naturejobs: <http://www.nature.com/naturejobs/science/jobs?q=china>
Science careers page: <http://jobs.sciencereers.org/jobs/life-sciences/china/>



In 2014, China's Three Gorges dam broke the world record for annual hydroelectric power production generating 98.8 billion kwh of electricity.



STEFFEN DUHM

>FOREIGN EMPLOYEE

>PROFESSOR SINCE 2012

>INSTITUTE OF FUNCTIONAL NANO AND SOFT MATERIALS (FUNSOM)
SOOCHOW UNIVERSITY

What encouraged you to move to China?

Before I took the position at Soochow University, I visited for two months and fell in love with Suzhou and its 2,500-year history. The old quarters of the city remind me of my hometown, Schwäbisch Hall, although the population of Suzhou is a hundred times larger and the history is four times longer. Before I came to China I lived in Japan for four years, but it is very difficult for foreigners to get positions above the postdoc level there, and in Germany there are almost no tenure-track positions.

How does your daily life contrast with living in the West?

My life is very Western in style. My apartment is to Western standards, there are some Western supermarkets, many pubs, and many, many German engineers here working at Bosch. If I wanted, I could live an almost entirely German life here, except for the air pollution and the Chinese internet restrictions.

How easy have you found it to make friends?

It has been somewhat difficult to make Chinese friends. The pubs in Suzhou are mainly frequented by expats. At work, most of my colleagues are very friendly, however, I get the impression most of them are not interested in becoming close friends with a foreigner – unless they've spent some time in English-speaking countries and are more familiar with Western customs.

How did your family react to your move?

My mother and my sister were very shocked when I told them. Although they were used to the fact that I was far away, they did not like the idea that I would live in China. But my mother has since visited and feels much more comfortable about it now.

Do you need to understand Mandarin to live in China?

I have not learned Mandarin. I can manage daily life like shopping, taking the bus or going to restaurants without

the language, but it requires some more planning. For example, I avoid restaurants without picture menus or taking a taxi spontaneously. In Japan, I started dating a Chinese woman, who has since moved here, and she helps me with practical things like finding an apartment or buying furniture.

Does the language barrier affect your working life?

At work Chinese – in principle – is not required. All faculty members of our college studied and worked abroad for some time and can speak English very well. All the teaching is supposed to be done in English. However, as all the students are Chinese, some of the professors switch to Chinese in their lectures. The communication with some of the students in my research group is a little difficult and they are de facto supervised by the Chinese associate professor in our group. The biggest problems are the written regulations and announcements. I get great help, but it is impossible for staff to translate all the documents for me, so they have to decide what is important. Sometimes their guesses are wrong, or reflect their opinion. Another problem is that most of the grant proposals have to be written in Chinese and consequently my proposals have to be translated, which does not enhance their quality.

What misconceptions do Westerners have about working as a scientist in China?

There are still many prejudices against Chinese science. I tried to convince some German friends to join FUNSOM at a faculty level, but they do not want to come to China because they think it might not help their future careers in Germany. For example, my PhD supervisor from Humboldt University was very, very skeptical and only came here the first time because we are friends. Now he is a chair professor of FUNSOM, visits here twice per year, and three students of FUNSOM will soon join his lab for their PhD. ■



DONGLAI FENG,
> EMPLOYER
> DIRECTOR OF THE STATE KEY
LABORATORY OF SURFACE PHYSICS,
FUDAN UNIVERSITY

Why should foreign researchers consider moving to China?

China is the fastest-growing place in terms of research funding, and the academic culture is transforming into a more healthy and creative one. At this time, there are great opportunities for scientists to practice their leadership skills and become pioneers. In the past ten years, all the basic infrastructure has been set up so labs are now as modern as those in many developed countries, with the added advantage that researchers can easily afford administrative support.

What cultural differences should those coming in from abroad be aware of?

The culture of science is still in the developmental stage here. Hierarchy is an important part of Chinese culture, so foreigners quickly learn to be tactful. At lunch, professors may talk more about politics than science. During meetings, academics often prefer not to make critical comments of their peers because they are afraid they might embarrass someone who could harm their career in the future. However, things are changing and becoming much more meritocratic.

**'THE ANTI-CORRUPTION
 CRACKDOWN HAS MADE
 PEOPLE MUCH MORE
 CAUTIOUS.'**

Are there any myths about working as a scientist in China that aren't true?

Corruption has been a problem in the past, but the recent anti-corruption crackdown by the Chinese government has made people much more cautious. There also used to be an instant financial bonus for getting published in major journals such as *Nature*. Now the emphasis is on the research and the 'paper bonus' culture is largely gone. ■

POSTDOCS IN CHINA



Fudan Daxue, Campus, Shanghai, China.

Good postdocs are the engines that drive much research and while Western universities are unable to offer places to all those who want postdoc work, in China there are plenty of positions but a lack of high quality candidates to fill them.

Poor pay and prospects have historically driven the country's cleverest PhD graduates overseas to gain experience in foreign labs. Without experience abroad, ambitious Chinese researchers stand little chance of obtaining tenure at home, and are paid poorly. Many potential postdoc candidates from abroad are put off by an outdated impression of Chinese science among ambitious graduates.

"Although science in China is growing tremendously in quality and quantity, the image of Chinese science makes it difficult for me to recruit foreign researchers, especially postdocs," explains physicist Steffen Duhm of Soochow University. "As my research is very fundamental and specialised, this is important, as it is difficult to find experts within China."

Many foreigners are concerned that taking a postdoc position in China could be detrimental to their careers. Back in 2011, China's minister of human resources and social security Yin Weimin pointed out that foreign postdocs account for 65% of the US postdoctoral workforce and 40% in the EU, but only 1% of China's postdocs.

"There has not been good track record for people coming out of China as postdocs to land good academic jobs," points out Ying Liu, a physicist who holds positions at both PennState University and Shanghai Jiao Tong University.

No-one is suggesting that the problem can be solved entirely with financial rewards, but they might be a good place to start, says Donglai Feng,

director of the State Key Laboratory of Surface Physics at Fudan University. His highest paid postdocs receive around \$34,000 [210,000RMB] per year with additional housing subsidies. He currently has 30 postdocs in his laboratory, 25% of whom are from foreign institutes.

**'THE GENERAL RESEARCH
 LEVEL HERE IS BECOMING AS
 HIGH AS IN OTHER COUNTRIES.'**

"The situation is changing dramatically because the general research level here is becoming as high as in other countries," says Feng. "My vision in the future is to recruit the best postdocs either from China or from other advanced countries. If there are not enough permanent positions for young people in these countries, China is a good option." ■



Between 2011 and 2013, China consumed more concrete than the US in the entire 20th century.

JAPAN

Being a team player is paramount in Japanese culture, and despite a shift in funding structures to short-term contracts, foreign researchers are attracted by cutting-edge science.

TON KOENE/ALAMY



Socialising with colleagues is a major part of teamwork required in the Japanese working culture.

When Louis Irving celebrated his marriage with a traditional Shinto wedding in the Japanese coastal town of Sendai in April, the first person to toast the couple wasn't an old friend or relative from home. It was Shinobu Satoh, his boss of five years.

"In Japan, your boss is considered one of your main sponsors in life, helping you to gain promotions and move up the food chain," says Irving, a biologist at the University of Tsukuba, north of Tokyo. "It was nice for my wife's family to hear that I am doing a good job at work and have a secure future."

Irving first arrived in Japan in 2003 as a PhD student from the University of Aberdeen with a keen interest in the enzyme RuBisCo, responsible for converting atmospheric carbon dioxide into fuel for plants.

"The best place for measuring the synthesis and degradation of RuBisCo was Tohoku University in Sendai," says Irving, who returned to Japan in 2007 and soon took up a postdoctoral fellowship through the Japanese Society for the Promotion of Science (JSPS).

Today he works as an assistant professor at Tsukuba, a position created under the government's ambitious Global 30 project, which aimed

to bring 300,000 international students to Japan. Last year, the government launched the Top Global Universities programme to encourage universities to become more international and to raise their global rankings through student and faculty exchanges, joint degrees and research collaborations, employment of more foreign staff and provision of more lectures in English.

**'YOUR BOSS IS
CONSIDERED ONE OF YOUR
MAIN SPONSORS IN LIFE.'**

The University of Tsukuba, for example, plans to increase its proportion of foreign researchers from the current 5% to 18% by 2023, but some believe that such aspirations are being undermined by a shift in strategic funding priorities.

"The 1980s was Japan-bashing time," says Atsushi Sunami, a specialist in science, technology and innovation policy at Tokyo's National Graduate Institute for Policy Studies. "The world criticized Japan for free-riding on fundamental ideas created in the United States and Europe, so the government decided to invest **PAGE 18 ►**



SHIGEO KOYASU

> **EMPLOYER**

> **EXECUTIVE DIRECTOR**

> **RIKEN**

What are the benefits of working at a national research institute in Japan as opposed to a university?

RIKEN is an environment in which researchers can collaborate across disciplines, unlike traditional Japanese university departments in which researchers have to ask permission just to talk to people from other labs. RIKEN also maintains excellent facilities. As an immunology researcher, for example, I need to use expensive cell sorters and DNA sequencers, and my centre has several of these.

Foreign researchers account for around 19% of researchers at RIKEN, higher than the national average but lower than the target of 30%. Having a heterogeneous population is good for science.

What are the differences from working culture in the West?

Japan does not have a culture of salary negotiation and they are much less flexible than in the US. People are just told what their salary is, based on age and experience. Performance schemes have been introduced, and if your performance is poorer than average, you may take a cut.

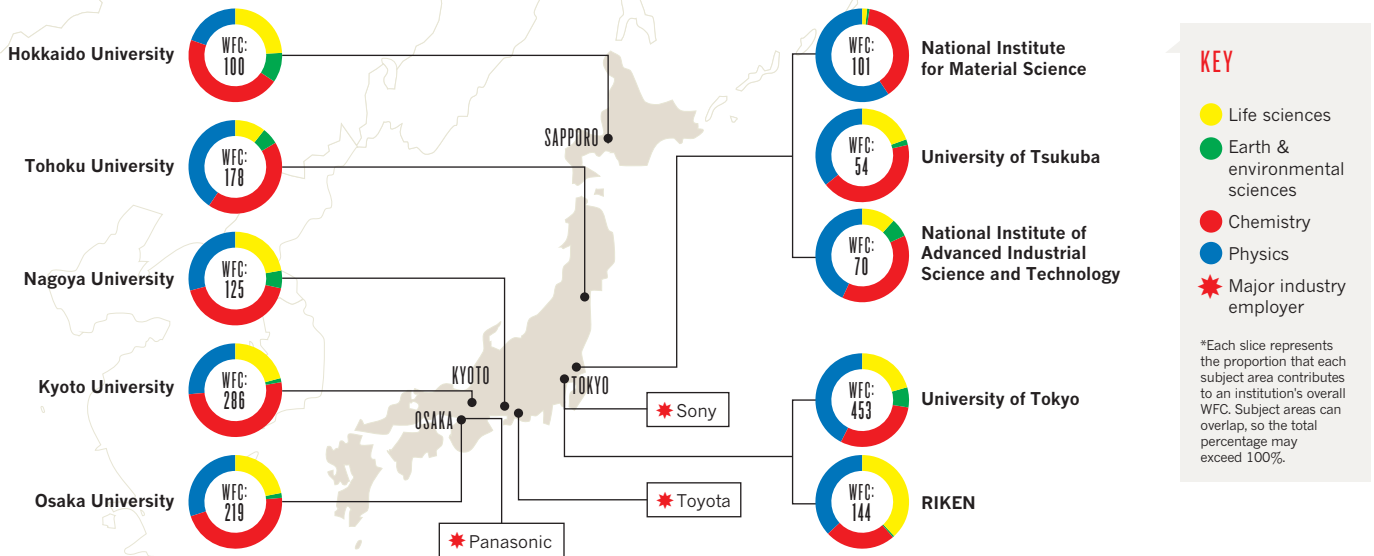
Is it important to speak Japanese?

Researchers do not need to speak Japanese to work in laboratories at RIKEN. We translate documents, hold meetings in English and assist researchers in banks and government offices. Foreign researchers can apply for funding in English. Difficulty for non-Japanese people becomes apparent when getting drivers' licences and social security numbers, for example. ■

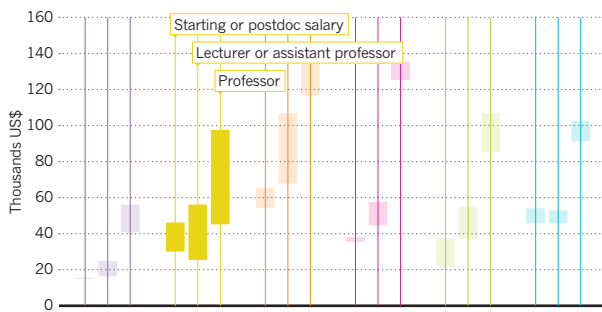
SOURCES: Funding over time: OECD; Spending per researcher: UNESCO; Research quality: Scopus/Nature Index; Salary ranges: Interviews & job ads; Reported well-being: Gallup World Poll

WHERE TO WORK

The below charts represent the research output included in the 2014 Nature Index for ten of Japan's leading institutions, and the contributions of different subjects, measured by weighted fractional count (WFC).

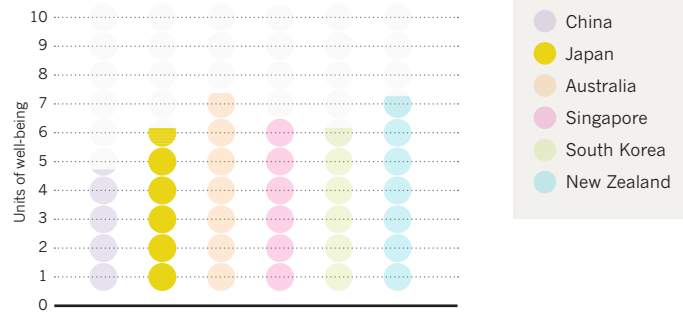


SALARIES



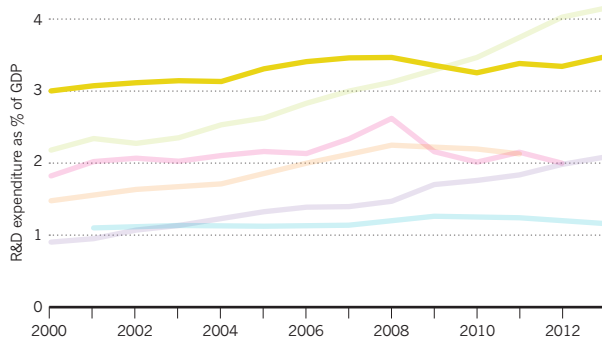
According to our ranges, which are on the basis of reported data, Japan has the smallest salary differences between academic staff grades.

REPORTED WELL-BEING



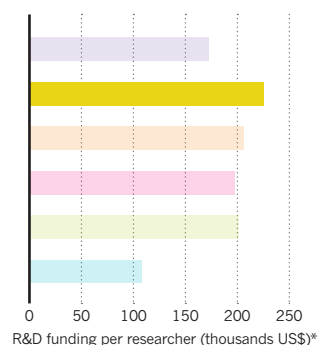
Only the Chinese report being less satisfied with their lives than the Japanese.

FUNDING OVER TIME



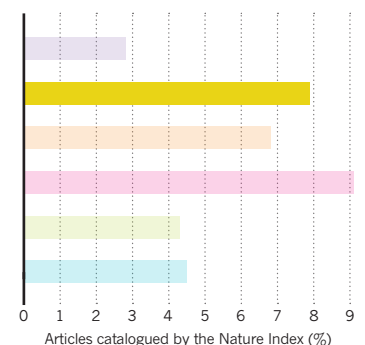
Over the plotted period, Japan has invested the highest proportion of GDP in R&D of the six countries. Investment is on the rise after a dip following the 2007-08 global economic slowdown.

SPENDING PER RESEARCHER



Japan spends the most per researcher.

RESEARCH QUALITY



Japan is second only to Singapore in the proportion of its research that is included in the Nature Index.

*Figures are normalized for purchasing power, and are the latest available for each country.

► more in basic research. But when the financial crisis hit, people became concerned about wasteful taxpayers' money."

By 2000, funding was being directed primarily at research with clear societal and economic returns. Universities now get a greater proportion of their income from government-commissioned research projects, and they recruit a larger number of researchers on temporary contracts rather than in more secure, long-term arrangements. "A lot of talented young people are looking for jobs elsewhere," says Sunami.

Still, for researchers such as Duncan McMillan on a two-year JSPS fellowship at the University of Tokyo, Japan offers the chance to work with pioneers in state-of-the-art laboratories. McMillan studies the process by which protons move between proteins in the membrane-bound respiratory chain to produce energy in cells.

His quest has seen him focus on microbiology in New Zealand, bioelectrochemistry in the United Kingdom and most recently, single-molecule microscopy in Germany. He was one of 240 foreign fellows selected under the JSPS scheme in 2014 — the top 10% of applicants.

"I realized that the technologies in Japan were above and beyond what I could access in Germany," says McMillan.

'LABS OPERATE AS ONE LARGE TEAM WITH STRICT HIERARCHIES'

Culturally, what sets labs apart in Japan is the way they operate as one large team with strict hierarchies, says McMillan. "You have to fit within that structure while maintaining your independence as a scientist — it can require a bit of give and take."

Team activities include dinner parties and football, but the real test of a team player in McMillan's laboratory takes place every Monday morning when undergrads and postdocs get together to clean. "Vacuuming office corners, mopping floors, removing rubbish and putting away glassware; it all happens incredibly quickly. Sometimes it can be difficult to find a job for yourself." ■



KOZ MASUMITSU

>EMPLOYER

>DEPUTY GENERAL MANAGER

>LSI MEDIENCE CORPORATION

What employment opportunities are there for foreign researchers in private companies in Japan?

As a clinical-testing company, most of our research involves developing cancer biomarkers and diagnostic reagents for commercial use. We have a few foreign nationals working in our research centres, but they were hired through the traditional recruitment system after graduating from a Japanese university. All of them can speak and write Japanese fluently.

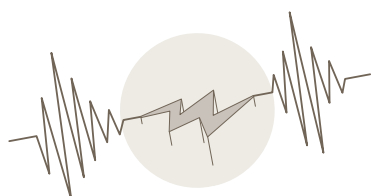
Overall, foreigners account for a small number of researchers in the private sector, but this varies by industry, with fewer opportunities in the chemicals and semiconductor industries, for example, than in the information technology sector. The government is discussing incentives for foreign companies to set up research and development centres based on models in Singapore, South Korea and China.

How does the research environment differ between the private and public sectors?

There is a lot of competition for funding in the public sector, which some may say is healthy, but it can also be a researcher's nightmare. Many young researchers, especially in highly innovative fields, are hired on fixed-term contracts for only two to three years. This makes it difficult to retain experienced staff — a major headache even for Nobel prizewinning stem-cell researcher Shinya Yamanaka. The funding environment is much less shaky in the private sector, once a company decides that a product is important for the business.

What's your advice to foreign researchers?

When in Rome, do as the Romans do. Cultural adjustment to daily life and the working environment in Japan can be challenging for foreigners. Japanese companies value teamwork and it can be difficult to figure out how decisions are made within the team structure. Sometimes patience is needed. The best practice is not to hurry — just wait and see. Build good relations with your colleagues, go out with them for dinner and drinks, and eventually you will begin to see results. ■



The Japan Meteorological Agency monitors 1,500 seismometers and 4,300 seismic-intensity meters, allowing it to report on the location and intensity of an earthquake within 90 seconds of its occurrence.

ENTRY REQUIREMENTS

Foreign nationals can apply for a Highly Skilled Foreign Professionals visa, which grants them (and a spouse) permission to work and live in Japan for five years, multiple re-entry, and consideration for permanent residence status after just five years (instead of the usual ten years). Employers must first issue a Certificate of Eligibility after seeing evidence of education level, expected salary and age and research achievements. Employees must include this certificate when they apply for their visa at their local Japanese embassy. The process takes between one and three months.

On entering Japan, new arrivals receive a residence card that they must carry at all times. They can leave and re-enter Japan for a maximum period of one year as many times as they wish.

ACADEMIC YEAR

Autumn term starts in October and ends in late February or March with a winter break from late December to early January. **Spring break** in March is followed by **spring term** which runs from April to late July. **Summer break** is in August.

OPPORTUNITIES & CONTACTS

The Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship for Overseas Researchers: <https://www.jsps.go.jp/english/e-fellow/postdoctoral.html>

JSPS Grants-in-Aid for Scientific Research:

<https://www.jsps.go.jp/english/e-grants/>
Japan Science and Technology Agency (JST) Strategic Basic Research Programs: <http://www.jst.go.jp/kisoken/en/about/index.html>

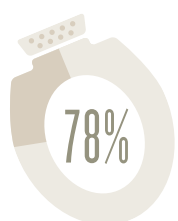
JST Impulsing Paradigm Change through Disruptive Technologies Program:

<http://www.jst.go.jp/impact/en/>

RIKEN Special Postdoctoral Researcher Program: <http://www.riken.jp/en/careers/programs/spdr/>

RIKEN International Program Associate:

<http://www.riken.jp/en/careers/programs/ipa>



78% of homes in Japan have an electronic toilet seat, more than those with a dishwasher, but fewer than have flat-screen televisions.

AUSTRALIA

Remote and beautiful, Australia is still a magnet for international researchers despite recent cuts to the budgets of some large scientific institutions and reduced grant opportunities.

NEAL PRITCHARD



CSIRO's ASKAP antennas at the Murchison Radio-astronomy Observatory in Western Australia.

What springs to mind when you picture Australia? Maybe it's the robust beauty of its coastline or the laid-back drawl of the locals. Perhaps you're not imagining an internationally competitive scientific powerhouse.

While the nation's 23.6 million people earn it a position of 51 in the world's population ranking, Australia reached number 11 for research-publication output in the global SCImago Journal & Country Rank in 2013, and last year came fourth in the *Scientific American Worldview's* biotechnology innovation scorecard.

Australia is a magnet for international researchers, with around 45% of scientists publishing research having come there from somewhere else, according to a survey published in *Nature Biotechnology* in 2012. One such researcher is Amanda Moffett, who moved from the US to become a post-doctoral research associate studying galaxy evolution at Perth's International Centre for Radio Astronomy Research.

"Astronomy is a small field, so everything is about how good a fit a place is for your interest," says Moffett. The pay for postdocs is also almost double what she would be receiving had she remained in the US, although much of this is accounted for by the increased cost of living

compared with other places, she points out.

Construction of the Square Kilometre Array, a radio telescope project 50 times more sensitive than any other radio-instrument, is due to begin in 2018 in Australia and South Africa. Australia is also home to the Australian Synchrotron particle accelerator, and one of the world's highest voltage heavy ion accelerators, as well as world-leading facilities in the emerging field of metabolomics.

**'IT'S NICE THAT IT'S A BIT SMALL
SO YOU GET TO KNOW A LOT OF
OTHERS IN THE SAME FIELD'**

For Moffett, Australia's relatively small population makes apparent the lack of women in science. "In the States the ratio [of men to women] is not that different, but the absolute size of the community here is much smaller so you feel like there are fewer higher-level women to look up to and have as mentors," Moffett says.

Sally Lavender, who moved to Australia from the UK with her partner after completing her PhD and is now a researcher at CSIRO Oceans and Atmosphere in Melbourne, has **PAGE 22 ►**



FRANK GANNON

> EMPLOYER

> DIRECTOR AND CEO

**> QIMR BERGHOFFER MEDICAL
RESEARCH INSTITUTE**

What's good about working in Australia?

There's a strong commitment to doing quality research and moving it through the system to have a clinical impact. It also has a great quality of life.

What are the challenges?

The world needs to know what you're doing, so researchers from here travel a lot to international scientific meetings. That is more necessary than if you were in mainland Europe and you could drop in on nearby labs for interest and profile raising. Funding is a challenge, but tell me a country where it isn't.

What are the main concerns for international job-seekers?

Firstly, the distance — but because of Skype it is easy to remain in contact. The second question that comes up is, why come here as opposed to elsewhere? That really is because of the quality of work. Then it gets down to visas. The 457 visa, you get that very readily, it's four years and you and your partner can work for that period. It can be renewed, and you can convert it into a longer term here or permanent.

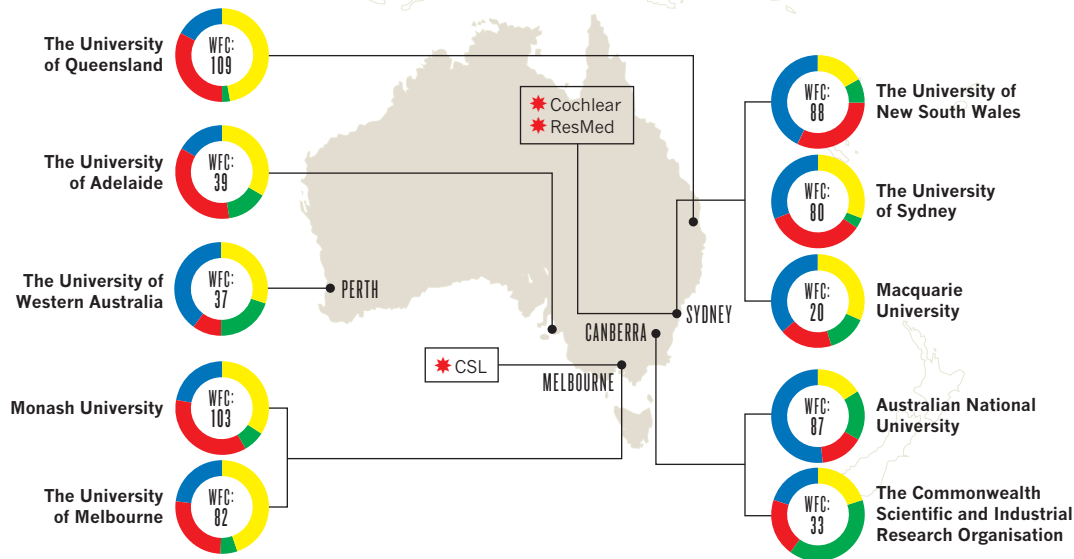
How can jobseekers improve chances?

Funding is one thing people should be looking for. If you get that, you've got your own salary, which is always very welcome, but you also then are able to say that you had that as an award. Organizations such as the European Molecular Biology Organization offer fellowships to anybody from Europe to go anywhere in the world, and that includes Australia. ■

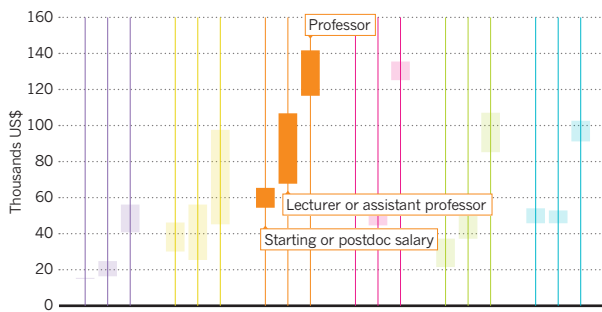
SOURCES: Funding over time: OECD; Spending per researcher: UNESCO; Research quality: Scopus/Nature Index; Salary ranges: Interviews & Job Ads; Reported well-being: Gallup World Poll

WHERE TO WORK

The below charts represent the research output included in the 2014 Nature Index for ten of Australia's leading institutions, and the contributions of different subjects, measured by weighted fractional count (WFC).

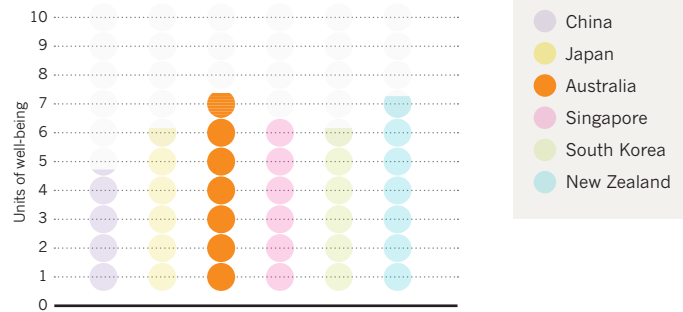


SALARIES



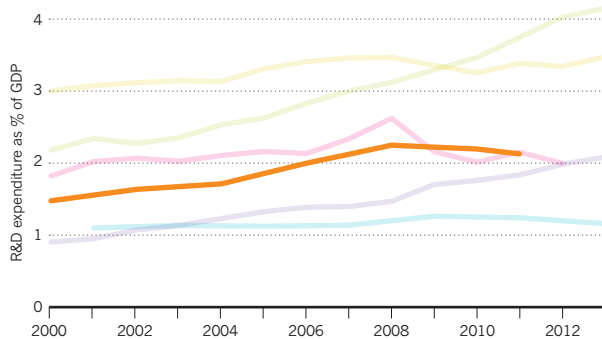
Australian researchers enjoy the highest salaries in the six countries, according to reported data, but the country is also among the most expensive places to live.

REPORTED WELL-BEING



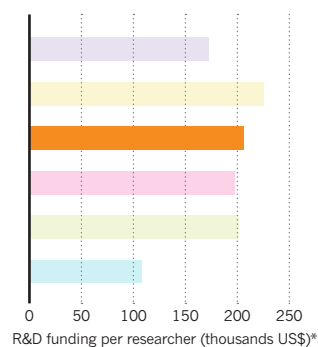
Australians report the highest levels of satisfaction with life among residents of the six countries.

FUNDING OVER TIME



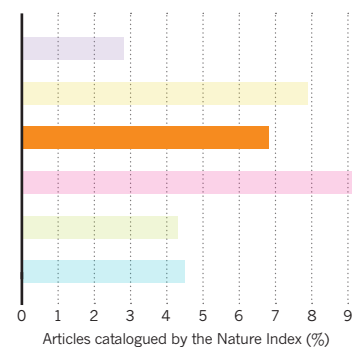
Some major scientific institutions in Australia have faced budget cuts as research and development (R&D) funding as a proportion of gross domestic product (GDP) has dipped from a high point of 2.2% in 2008.

SPENDING PER RESEARCHER



Australia invests the second-highest amount per researcher, after Japan.

RESEARCH QUALITY



The proportion of Australia's research output included in the Nature Index is a little over average for the six countries.

*Figures are normalized for purchasing power, and are the latest available for each country.

► come to appreciate the country's scientific microcosm. "It's nice that it's a bit small, so you get to know a lot of other scientists that are in the same field as you, [and see] the same people at domestic conferences," she says. Having recently had her first child, Lavender says her experience of maternity leave and the public health system was "brilliant".

Less positive, Lavender says, is that funding constraints are tightening, and not just for climate-related research. Indeed, there have been major ups and downs in the funding of science in recent years, with cuts to some of Australia's biggest scientific institutions including the CSIRO and the Cooperative Research Centres — research consortia of industry, university and government organizations designed to boost economic growth.

The axe has also fallen on the Australian Research Council's Future Fellowships, which used to support 200 mid-career researchers including overseas applicants for four years. That has recently been cut to just 50 fellowships, with preference given to Australian applicants.

On the flip side, the Federal Government has this year committed to funding an AUD\$20 billion Medical Research Future Fund, and thrown a last-minute financial lifeline to the Australian Synchrotron.

"The really high-quality science is still getting funded," says James Murphy, who moved to Australia from New Zealand to undertake a PhD in biological chemistry, and is now a group leader at the Walter and Eliza Hall Institute of Medical Research in Melbourne. He has been involved in assessing grant applications for the National Health and Medical Research Council, and says that although there has been a recent decline in the number of grants offered, "there are still initiatives and a diversity of funds to which you can apply".

Murphy recommends sourcing funds internationally before coming to Australia. "It's a fine training ground; facilities-wise it's world class," he says, "but, bring as much money as you can from a foreign body and use that to fund your work and your salary while you're here." ■



JOHN HEASMAN

> **EMPLOYER**

> **PRINCIPAL RESEARCH ENGINEER**

> **COCHLEAR LTD**

What does Australia have going for it as a place to live and work?

For a small nation, real benefits come from focusing on areas where you have strengths. Australia is leading the world in some areas of research innovation, such as applied clinical care, cancer treatment, and vaccines. The climate and the lifestyle are huge draw, particularly in Sydney and Melbourne. It's also a very multicultural society now, which has really facilitated people moving from Europe, the US and other countries. Cochlear has people from 71 nationalities in its Sydney headquarters.

Are there any downsides?

The geographical separation is a challenge. Though teleconferencing and the Internet have revolutionized communication, it's not the same as working face to face.

What are the main questions international jobseekers have?

Generally, overseas applicants first ask about who to approach for opportunities. Our advice is to look for Centres of Excellence with Australian Research Council funding. Generally they're groups that have good academic and industrial ties, and collaborate very closely. The main aim of those relationships is to work on cutting-edge research and make sure it is applied and relevant to commercial industry. We're big believers in the power of basic research, but the real power is translating that into clinical practice. If you're interested in working in private industry, you're best to approach some of the bigger players, such as Cochlear and ResMed which have strong research and development programmes.

What do foreign job seekers bring to Australia?

Through dealing with their supervisors and the running of their research, they establish ties with other collaborators, so often bring connections with them. The other advantage is knowledge transfer. We're isolated geographically and it's important to not just rely on publications, but have human interactions. ■



Australian-born physicist Sir William Lawrence Bragg and his father Sir William Henry Bragg who jointly won a Nobel Prize in 1915 are the only father-son team to have ever been awarded the honour.

The document most commonly used by researchers coming to Australia is the 457 visa. This is a temporary visa enabling skilled workers to come with their dependents for employment in their chosen field for up to four years. They must be either sponsored by an employer or be a good fit for an available position. These visas are generally processed within 90 days of submission and currently cost AUD\$1035 (US\$800). Another option is the non-sponsored 189 skilled independent visa, which costs \$3520. English-speaking qualified applicants under the age of 50, who could help fill the country's skill shortages, are assessed under a points-based system.

Most Australian universities have two terms, the first of which begins in late February or early March and runs until May or June with a break in April. The **mid-year recess** is late June until late July. The **second term** runs from then until late November, with a break in late September or early October.

OPPORTUNITIES & CONTACTS

The Australian Research Council's Discovery Projects – Discovery International Awards:

http://www.arc.gov.au/general/international_researchers.htm

The Australian Research Council's Australian Laureate Fellowships: http://www.arc.gov.au/nccgp/laureate/laureate_default.htm

The Group of Eight (G08) European Fellowship scheme: <https://go8.edu.au/article/go8-media-release-european-fellows-2015>

The Endeavour Scholarships and Fellowships: <https://internationaleducation.gov.au/endeavour%20program/scholarships-and-fellowships/international-applicants/pages/international-applicants.aspx>

The Connecting Australian European Science & Innovation Excellence programme: <http://www.caesie.org/>



Australia's mainland coastline is 35,000 km long, with more than 10,000 beaches.

SINGAPORE

A hub for proactive and dynamic professionals, funding and salary levels in this ambitious city-state make the opportunities on offer hard to resist for many.

HUTTON AND CROW



Nanyang Technological University's new Learning Hub, which features state-of-the-art classrooms.

It's such an international, positive environment," says Sara Sandin. A Swedish assistant professor in the school of biological sciences at Nanyang Technological University who came to Singapore in 2012 to set up an electron microscopy laboratory, Sandin speaks highly of her adopted home. "Research budget cuts elsewhere make it hard to maintain a positive and creative working environment. You don't have that in Singapore."

Between 2011 and 2015, 16.1 billion Singapore dollars (US \$12 billion) was invested in science and research by the island city-state, a 20% increase on the previous five-year period. "The long-term goal is to turn Singapore into a world-renowned research hub," says Terence Ong, of Contact Singapore, a government agency that seeks to attract people to live and work on the island. "Singapore is committed to keeping up momentum for money going into research." In a world of static or shrinking budgets, this makes Singapore hard to resist for many.

Science is increasingly carried out in collaborations between academia and industry. In 2014, for example, global IT giant Fujitsu set up the Centre for Excellence on Sustainable Urbanisation in collaboration with the Singapore Management University and the government's

Agency for Science, Technology and Research (A*STAR), which plays a leading part in encouraging these combined efforts. "We believe that this collaboration will bring a lot of value, some that we cannot even predict," says Rio Yamaura, vice president of the New Solutions Business Division at Fujitsu.

Although Singapore entered a period of deflation at the end of last year, high inflation between 2010 and 2014 made it one of the world's most expensive cities to live in. A one-bed apartment in the city centre costs anywhere between SGD2500 to more than SGD8000 per month. But high salaries and low income tax rates (maximum 20%) allow for a comfortable standard of living.

"We advise incoming scientists to use our cost-of-living calculator before arriving so they can prepare themselves," says Ong. There is assistance through housing allowances and some on-campus housing offered by universities to researchers for, on average, one-third of the market rate. ■



Despite having no oil, Singapore is the world's largest manufacturer of offshore oil-rig platforms.

ENTRY REQUIREMENTS

Once a foreign scientist has accepted a job offer, they need an Employment Pass from the Ministry of Manpower. Their employer must do the paperwork once it has received proof of education documents and CV

details, and the employee presents it to the ministry. Those wanting to explore work opportunities in Singapore can apply for a Personalised Employment Pass, which allows professionals to remain in Singapore for six months.



BERTIL ANDERSSON

> **EMPLOYER**

> **PRESIDENT**

> **NANYANG TECHNOLOGICAL UNIVERSITY**

How has Singapore changed since you arrived?

When I came here, I joked that Europeans talk too much, but Singaporeans act. And the joke became my reality. There was good money and quick decisions, so things grew rapidly. The investment in academia is monumental if you look at the size of the country, and that has all happened in the last 10-15 years. I've recruited a lot of professors in science and engineering, and very few leave Singapore. People love it, professionally and personally.

What are the challenges for foreign scientists?

I call it Asia-lite: a mix of East and West, so Europeans and Asians feel at home. It's very humid, so you have to live in an air-conditioned box. From a professional standpoint, this part of Asia has a lot of corruption. The government has fought it with heavy bureaucracy. You may need 11 signatures to buy one lab chemical.

Are there any cultural differences foreigners should be aware of?

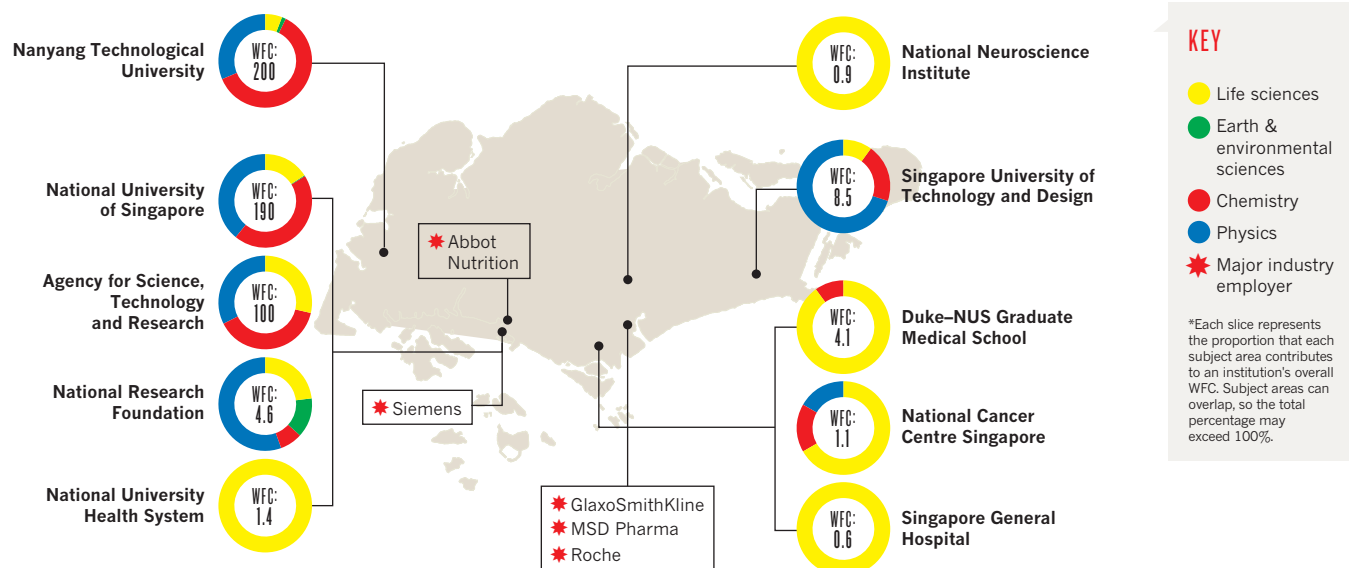
People don't always say what they think. Singaporeans don't want to lose face. If I tell my professors that they did something wrong, they take it very badly, so you have to be diplomatic. It's a secular society and religion is private. Singaporeans appreciate harmony and the government plays a watchful role in maintaining this.

How can foreigners prepare?

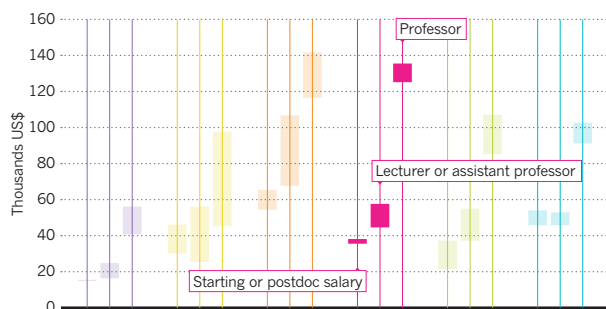
Have some money saved and a credit card! It's not cheap to live here. You need to be a proficient English speaker. ■

WHERE TO WORK

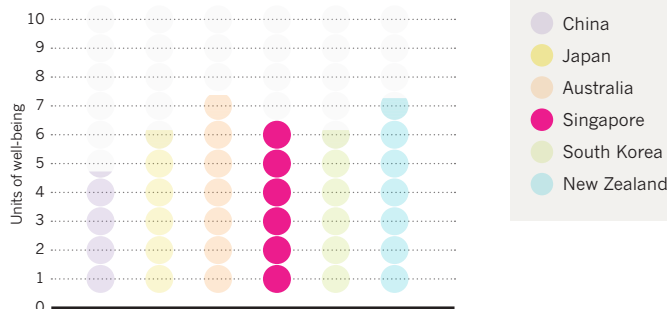
The below charts represent the research output included in the 2014 Nature Index for ten of Singapore's leading institutions, and the contributions of different subjects, measured by weighted fractional count (WFC).



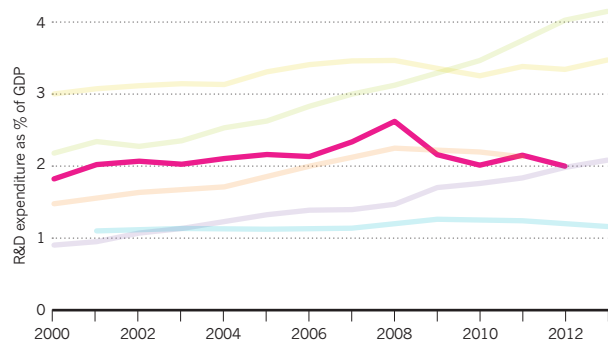
SALARIES



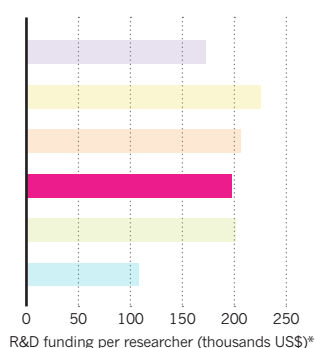
REPORTED WELL-BEING



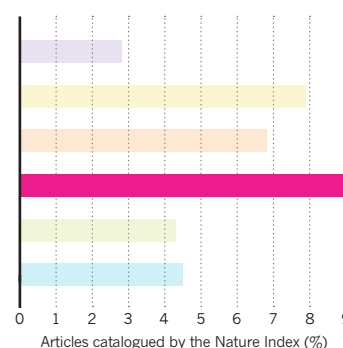
FUNDING OVER TIME



SPENDING PER RESEARCHER



RESEARCH QUALITY



*Figures are normalized for purchasing power, and are the latest available for each country.

SOUTH KOREA

Focused on recruiting overseas researchers and encouraging basic science, this Asian tech hub is using research to drive development

SEONGJUN CHO/BLOOMBERG/GETTY IMAGES



Among the fruits of R&D investment are companies like LG Electronics, whose headquarters are in Seoul.

When it comes to science, South Korea is an ambitious player. It invests more of its GDP in research and development than any other developed nation except Israel. This has helped propel the country's rapid rise from war-torn nation to G20 summit host and home of powerhouses like Samsung and LG.

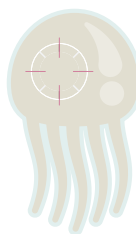
But only within the past decade have universities begun to employ more scientists from abroad, boosted by a government programme established in 2009 that committed around \$700 million to help universities bring talent from abroad. This recruitment drive is not without complication, not least because of cultural and language barriers. Foreigners often feel disadvantaged seeking funding as many agencies require applications to be submitted in Korean.

The attempt to attract more foreign recruits coincides with efforts to boost basic research, after decades of focusing on applied research and economic growth. President Park Geun-hye has pledged to ease the path for tech start-ups, limit the influence of the tech giants and to kickstart a "creative economy". Halfway through her five-year term, the results have been mixed. The Institute for Basic Science, a research centre founded in Daejeon in 2011 aimed at rivalling

Japan's RIKEN, has since opened other centres, but scientists say that grant pools are spread thin, and levels of individual grants are flat or falling.

On the positive side, junior researchers in South Korea enjoy relatively great freedom and are not expected or obliged to work on projects run by senior academics.

Foreigners are often surprised by the array of equipment at their disposal. "For biology, there is everything we need in Korea," says Eric di Luccio, a structural biologist from France who joined Kyungpook National University in Daegu in 2010. That includes national facilities such as the Pohang Accelerator Laboratory that are not yet on the radar of researchers who might trek to Japan or Taiwan for instrument time. "There is a lot of infrastructure and research going on, and it's not widely recognized yet," says di Luccio. ■



South Korea's Jellyfish Elimination Robotic Swarm (JEROS) provides an automated solution to problematic jellyfish blooms. Robots hunt in packs, each capable of pulverizing 400 kg of jellyfish every hour.

Thanks to South Korea's push to make its scientific workforce more international, the government freely approves work visas for researchers, though anecdotal reports suggest there can be hold-ups at the university or department level if institutions are not accustomed to employing foreigners and are unfamiliar with the required documentation. These

visas need to be renewed annually. Foreign nationals whose parents or grandparents emigrated from South Korea — or who were once citizens themselves, including those adopted abroad — are eligible for an Overseas Korean F-4 residence visa. This grants nearly all the benefits of citizenship except voting rights, and is not specific to a particular employer.



HAKIM DJABALLAH
> EMPLOYER
> CHIEF EXECUTIVE OFFICER
> INSTITUT PASTEUR KOREA

Are there good opportunities for postdocs in Korea?

Coming here for three years or so is a great opportunity for a postdoc considering it as a little bit of a break to do some teaching, enjoy life in Korea and explore Asia. You may struggle in terms of publications unless you're at Seoul National University. Korea has been struggling with publishing in high-profile journals because editors and reviewers get more sceptical with the things that they send because of past fraud. I think many editors have had their fingers burned, and they often consider our researchers guilty until proved innocent.

What are the main obstacles for foreign scientists in Korea?

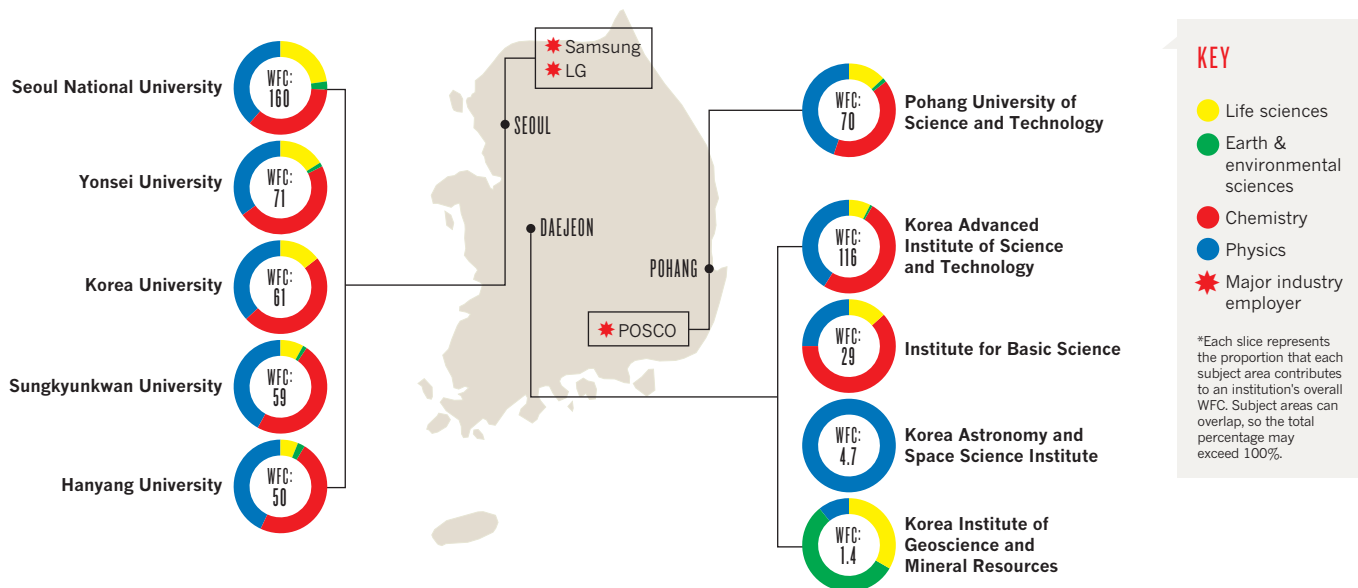
The lack of transparency and fairness of the granting and peer review systems. Foreign researchers in Korea suffer a lot because of the old boy network. It's often a case of, 'We went to kindergarten together, so I'll fund you.' You would never expect foreign scholars to be required to write grant applications and to have to defend their proposal in Korean, yet often that is expected. My researchers here suffer from that. We feel discriminated against.

What advice would you give to scientists coming to Korea?

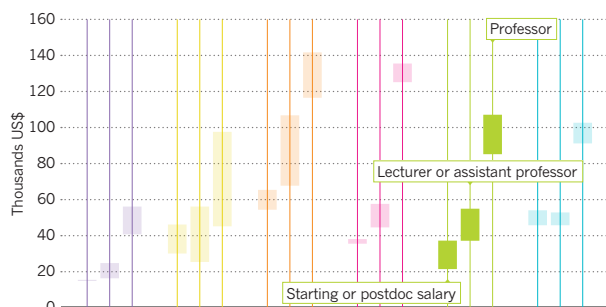
Negotiate everything in writing before you get here. The Koreans have a tendency to say, 'It's OK, we'll take care of it once you get here.' Scientists need to get agreements on everything from salary to health insurance, all the way to if they are married and want to bring their family, how they're going to do it. ■

WHERE TO WORK

The below charts represent the research output included in the 2014 Nature Index for ten of South Korea's leading institutions, and the contributions of different subjects, measured by weighted fractional count (WFC).

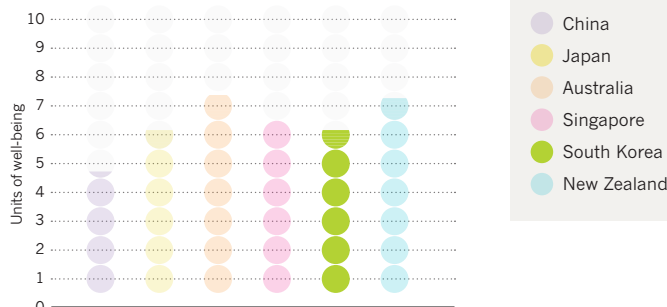


SALARIES



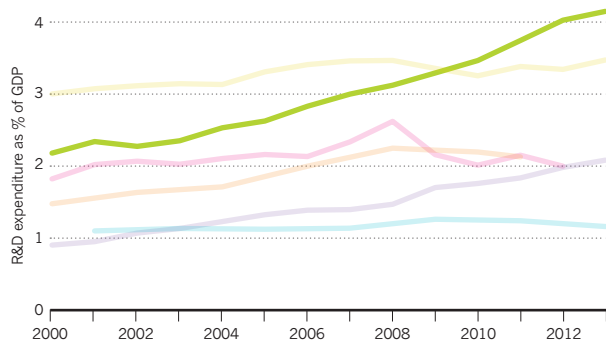
Postdocs are relatively badly paid and professors better paid compared to the same grades elsewhere in the region, according to our ranges based on reported data.

REPORTED WELL-BEING



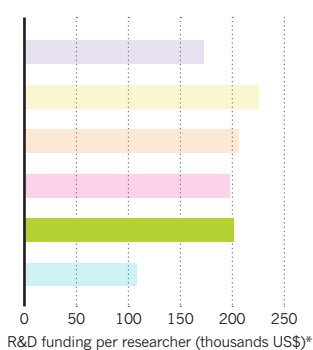
Life satisfaction among South Koreans is just below average for the six countries.

FUNDING OVER TIME



South Korea has roughly doubled its proportion of gross domestic product (GDP) spent on research and development (R&D) in the past decade.

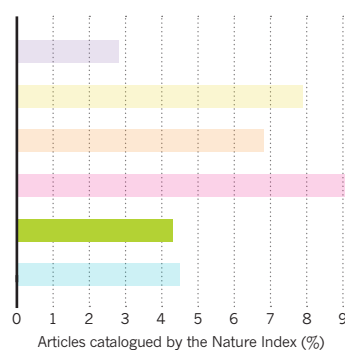
SPENDING PER RESEARCHER



Despite investing the highest proportion of its GDP in R&D, South Korea spends only the third most in the region per researcher.

*Figures are normalized for purchasing power, and are the latest available for each country.

RESEARCH QUALITY



Only China scores lower than South Korea in the proportion of scientific papers published in Nature Index journals.

NEW ZEALAND

A natural laboratory like no other, the small country is renowned for its Earth science and agricultural research. It has a multi-cultural environment and is proud of its collaborativeness.

ATANAMIR/WIKIMEDIA/CC BY-SA 4.0



Aoraki/Mount Cook is one of the many wonders attracting researchers to New Zealand.

Perennially snow-capped and tantalizingly steep, Aoraki/Mount Cook is a magnet for elite mountaineers. New Zealand's highest peak also symbolizes the spectacular natural environment that draws many to study and work in this South Pacific island nation.

The geologically unstable juncture of two tectonic plates, which has pushed Aoraki to 3,724 metres, regularly shakes the country and underpins much of the science at which New Zealand excels. It was the vantage point for studying geophysics and Earth sciences that attracted Pegah Faegh-Lashgari from Iran to do her PhD at the Victoria University of Wellington in 2012. "The university staff were thoughtful and things like accommodation were all sorted. The only negative has been the distance from Europe and my home country," says Faegh-Lashgari.

Faegh-Lashgari would consider staying on after her PhD, but expresses a widespread concern: the prospects for postdocs and for pursuing an academic career are limited. "New Zealand has very few postdocs," says Siouxsie Wiles, a University of Auckland microbiologist who relocated from London six years ago.

Yet it may be that New Zealand's small population has helped foster its ingenuity, known as

the 'number 8 wire tradition' — a reference to a gauge of fencing wire that has been adapted for uses beyond its original purpose. It is an attribute that has helped punch above its weight in the Nobel laureate-stakes. New Zealand has three; Ernest Rutherford, for chemistry (1908), Maurice Wilkins, physiology or medicine (1962) and Alan MacDiarmid, chemistry (2000).

About one quarter of New Zealand's R&D spending is accounted for by seven government-owned companies called Crown Research Institutes (CRIs). The two largest are focused on plant and food research, and agricultural research. Fonterra, a giant dairy-farming cooperative is responsible for almost one third of global dairy exports. With some 400 employees, it is the country's largest private organization doing research. New Zealand spends 1.14% of its GDP on R&D — much less than the OECD average of 2.4%. There is widespread support for government efforts to lift that to 2%. ■



New Zealand-born Ernest Rutherford, recipient of the 1908 Nobel Prize in Chemistry, discovered the proton and was the first person to split an atom.

ENTRY REQUIREMENTS

New Zealand has been proactive in developing its economy since the global financial crisis, but has many specialist skills shortages. It is recruiting overseas candidates to fill positions in fields such as medicine, engineering, information and communications technology and agriculture and forestry. Foreign applicants can apply for jobs before they have a visa, and

employers usually assist in obtaining one. For those with scientific qualifications, this will be the Skilled Migrant Visa, for which people aged 20-55, who are proficient in English and who meet certain health requirements are eligible. To work in some areas, notably engineering and medical and health professions, there is also a legal requirement to be registered with the appropriate professional body.



MICHAEL MCWILLIAMS

> EMPLOYER

> CHIEF EXECUTIVE

> GNS SCIENCE

What is particularly attractive about working as a scientist in New Zealand?

From the Earth-sciences perspective, it's a fantastic natural laboratory. We have all manner of active features that represent both significant geological hazards to the country and great research opportunities. New Zealand is a truly amazing geological sandpit, and scientists want to come. It's a great place for collaboration.

Is the geographical isolation a problem for career development?

None of us is isolated; we have the tools to collaborate as long as we're flexible about time zones. Europe and North America are less than a day away, Asia is closer, and at any time 15% of my scientists are overseas.

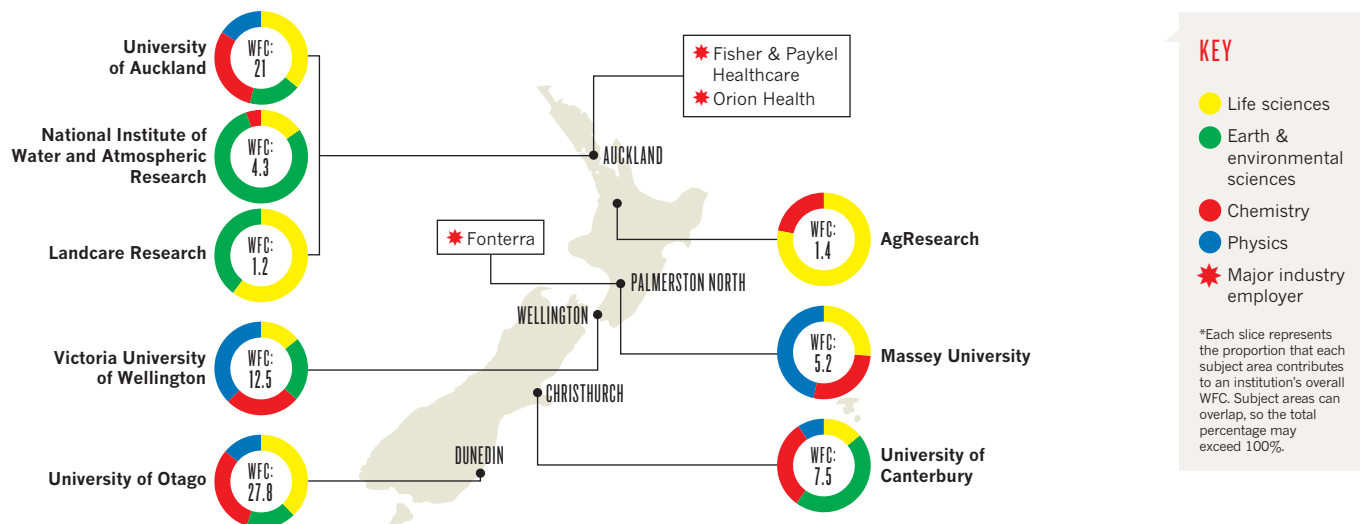
What cultural aspects should those coming to work from abroad be aware of?

New Zealand is a multicultural nation like many others, so a cosmopolitan approach is a benefit. A desire to learn about and respect Māori culture is important.

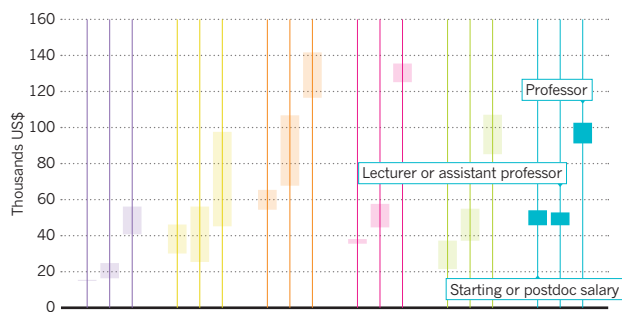
GNS Science is one of the seven Crown Research Institutes (CRIs), government-owned companies that do one quarter of R&D in New Zealand. How does this unusual structure affect the work they do? The CRI business model, in which research is funded by government and private sources and administered on corporate principles, is unusual, but not unique. Australia's CSIRO, the Fraunhofer Institutes in Germany, and the Dutch TNO are all similar. The principal advantage is the ability to make clear, visible and valuable contributions of national benefit. ■

WHERE TO WORK

The below charts represent the research output included in the 2014 Nature Index for eight of New Zealand's leading institutions, and the contributions of different subjects, measured by weighted fractional count (WFC).

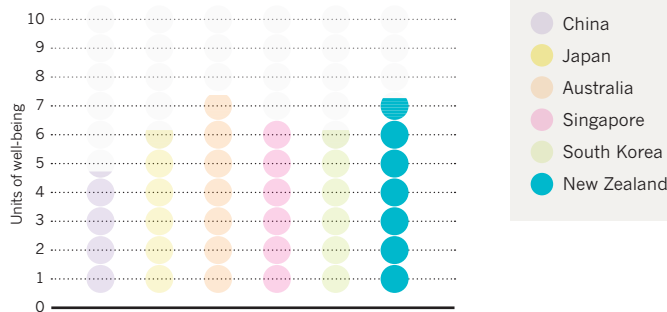


SALARIES



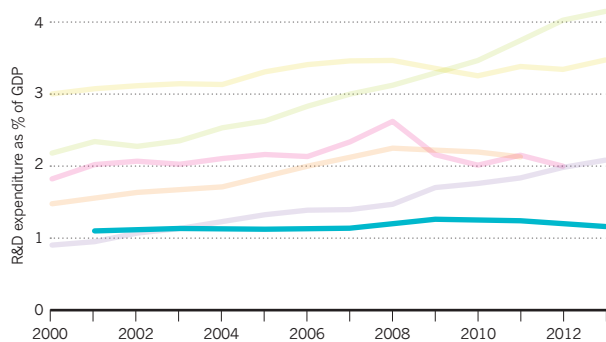
Salaries are neither particularly high nor low compared to the other six profiled Asia-Pacific countries, although postdocs do relatively well according to the ranges calculated on the basis of reported data.

REPORTED WELL-BEING



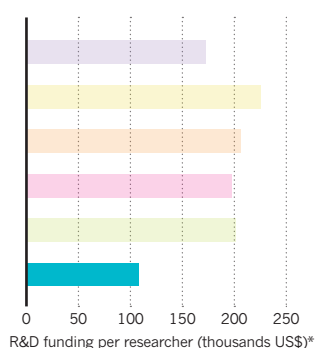
New Zealanders report high life satisfaction, just a fraction below that of the Australians.

FUNDING OVER TIME



Although others in the Asia-Pacific region have seen large rises in research funding, the proportion of GDP spent on R&D has remained consistently low in New Zealand.

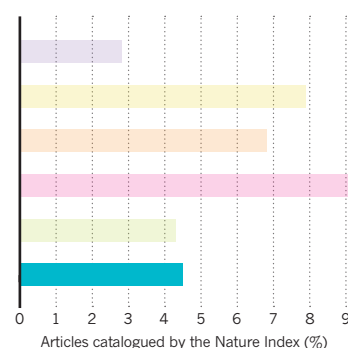
SPENDING PER RESEARCHER



New Zealand spends the least per researcher and less than half of that spent by regional leader Japan.

*Figures are normalized for purchasing power, and are the latest available for each country.

RESEARCH QUALITY



Measured by the proportion of its scientific papers included in Nature Index journals, New Zealand's research quality is below the average of the six countries.